Performance Metrics:
An Operational Impact Evaluation Plan

Version 1.0

August 12, 1999

Free Flight Phase One (FFP1)
1500 K Street, NW
5th Floor
Washington, DC 20005
(202) 220-3300

12 August 1999

This Metrics Plan represents a new approach by the FAA to proactively measure the performance impacts of new systems. The ideas presented reflect a collaborative effort between the FAA and the aviation industry. The FFP1 Metrics Team will use this plan as a roadmap for the evaluation of FFP1 benefits.

We would like to thank the numerous organizations and individuals who provided valuable inputs and critique.

Sincerely,

David A. Knorr

FFP1 Team Lead – Performance Metrics

Preface

Free Flight Phase One (FFP1) is a new juncture for the FAA. FFP1 capabilities were recommended by users of the National Airspace System (NAS) and air traffic controllers. These stakeholders will continue to play a role in the evaluation of FFP1's success. In the past, benefit estimates for new air traffic control systems were often made independent of airspace users (airlines, etc.) and minimal quantitative data was collected to determine if benefits were actually delivered. With FFP1, the sites for the new equipment are being monitored a year in advance of installation; and each site will be evaluated for performance impacts over the first year the new capabilities are in place.

Performance impacts are captured by metrics such as "arrival/departure rate at peak periods" or "time to complete a flight segment". Baseline data for specific metrics will isolate performance changes driven by FFP1 from performance changes driven by varying conditions such as weather, runway configuration, or traffic volume.

FFP1 has established a Metrics Team to interface with stakeholders to determine appropriate performance measures and evaluation methodologies. This evaluation plan reflects a collaborative effort between the FAA and the aviation industry. The process of establishing metrics clarifies the benefits sought by users. Execution of the FFP1 Metrics Plan will provide better information which leads to better decision making.

Table of Contents

Section	Page
1. Introduction	**1-1**
1.1 Background	1-1
1.2 Purpose and Scope of Plan and Evaluation	1-3
1.3 Data Requirements	1-4
1.4 Reporting Requirements	1-5
1.5 Organization	1-5
2. FFP1 Program Overview	**2-1**
2.1 FFP1 Capabilities	2-1
2.1.1 User Request Evaluation Tool	2-1
2.1.2 Traffic Management Advisor – Single Center	2-1
2.1.3 Passive Final Approach Spacing Tool	2-2
2.1.4 CDM: Enhanced Ground Delay Program	2-2
2.1.5 CDM: NAS Status Information	2-3
2.1.6 CDM: Collaborative Routing	2-3
2.1.7 Surface Movement Advisor	2-3
2.2 FFP1 Implementation Sites	2-4
3. Evaluation Overview	**3-1**
3.1 Evaluation Schedule	3-1
3.2 Categorizing Airspace User Objectives	3-3
3.2.1 Safety	3-5
3.2.2 User Access	3-6
3.2.3 Delay/Efficiency	3-6
3.2.4 Predictability	3-7
3.2.5 Flexibility	3-8
3.3 Operational Impacts and Performance Metrics	3-8
3.4 Evaluation Methodology	3-14
3.4.1 Data Collection	3-14
3.4.2 Data Reduction and Analysis	3-14
3.5 Economic Valuation	3-15
3.6 Environmental Impacts	3-16
4. Plan for Operational Evaluation of FFP1 Capabilities	**4-1**
4.1 User Request Evaluation Tool	4-1
4.1.1 Evaluation Overview	4-1
4.1.2 Performance Metrics	4-2

4.1.3 Evaluation Schedule 4-14
4.1.4 Data Collection 4-15
4.1.5 Evaluation Issues/Concerns 4-19
4.2 Traffic Management Advisor – Single Center 4-19
4.2.1 Evaluation Overview 4-19
4.2.2 Performance Metrics 4-21
4.2.3 Evaluation Schedule 4-26
4.2.4 Data Collection 4-27
4.2.5 Evaluation Issues/Concerns 4-30
4.3 Passive Final Approach Spacing Tool 4-32
4.3.1 Evaluation Overview 4-32
4.3.2 Performance Metrics 4-33
4.3.3 Evaluation Schedule 4-38
4.3.4 Data Collection 4-39
4.3.5 Evaluation Issues/Concerns 4-41
4.4 CDM: Enhanced Ground Delay Program 4-42
4.4.1 Evaluation Overview 4-43
4.4.2 Performance Metrics 4-45
4.4.3 Evaluation Schedule 4-53
4.4.4 Data Collection 4-54
4.4.5 Evaluation Issues/Concerns 4-55
4.5 CDM: NAS Status Information 4-55
4.5.1 Evaluation Overview 4-55
4.5.2 Performance Metrics 4-56
4.6 CDM: Collaborative Routing 4-56
4.6.1 Evaluation Overview 4-56
4.6.2 Performance Metrics 4-57
4.7 Surface Movement Advisor 4-58
4.7.1 Evaluation Overview 4-58
4.7.2 Performance Metrics 4-59
4.7.3 Evaluation Schedule 4-62
4.7.4 Data Collection 4-63
4.7.5 Evaluation Issues/Concerns 4-64

5. Evaluation Roles and Responsibilities 5-1
5.1 Free Flight Phase One Metrics Team 5-1
5.2 FFP1 Users and Service Providers 5-1

List of References RE-1

Appendix A. RTCA FFP1 Core Performance Metrics A-ii

Appendix B. Changes in Metrics from Previous Versions of the Metrics Plan B-1
B.1 User Request Evaluation Tool B-1
B.2 Traffic Management Advisor – Single Center B-1
B.3 Passive Final Approach Spacing Tool B-2
B.4 CDM Enhanced Ground Delay Program B-3
B.5 Surface Movement Advisor B-5

Glossary GL-1

List of Figures

Figure	Page
Figure 3-1. Integrated Evaluation Schedule (Prototypes)	3-1
Figure 3-2. Integrated Evaluation Schedule (Current and Future Deployments)	3-2
Figure 3-3. Stages of Deployment	3-3
Figure 4-1. URET Evaluation Schedule	4-15
Figure 4-2. Notional Terminal Airspace	4-23
Figure 4-3. TMA Evaluation Schedule	4-27
Figure 4-4. DFW TRACON Airspace and Meter Fixes	4-35
Figure 4-5. Representative Arrival Rates, DFW	4-37
Figure 4-6. pFAST Evaluation Schedule	4-39
Figure 4-7. Flying Time Metric – Feb 16-17, 1999; GDP-E @ SFO	4-47
Figure 4-8. Cumulative Compression Benefits (Jan 20, 1998 – Mar 17, 1999)	4-48
Figure 4-9. EDCT Compliance History	4-51
Figure 4-10. Percentage of Cancellations of GDPs from 9/8 to 12/31	4-52
Figure 4-11. SMA Taxi-Out Time Comparison, Atlanta Hartsfield International	4-61
Figure 4-12. SMA Evaluation Schedule	4-63

List of Tables

Table	Page
1-1. Description of FFP1 Capabilities from Addendum 1	1-2
2-1. Sites for FFP1 Implementation	2-5
3-1. FFP1 Expected Operational Impacts and Performance Metrics	3-11
4-1. URET Performance Metrics	4-3
4-2. URET Metrics, Data Elements, and Data Sources	4-18
4-3. TMA Performance Metrics	4-21
4-4. TMA Metrics, Data Elements, and Data Sources	4-29
4-5. pFAST Performance Metrics	4-33
4-6. pFAST Metrics, Data Elements, and Data Sources	4-40
4-7. GDP-E Performance Metrics	4-44
4-8. Sample EDCT Compliance Data	4-51
4-9. Top GDP Airports and Average Number of GDPs(1995-1998)	4-54
4-10. CR Performance Metrics	4-57
4-11. SMA Performance Metrics	4-59
4-12. SMA Metrics, Data Elements, and Data Sources	4-64
A-1. RTCA FFP1 Core Performance Metrics	A-1

Section 1

Introduction

This *Performance Metrics: An Operational Impact Evaluation Plan* (hereafter *Metrics Plan*) presents the performance metrics and evaluation methodology designed to measure the operational impacts of the Free Flight Phase One (FFP1) capabilities.

1.1 Background

FFP1 is an initial step in the evolution to *Free Flight*, which the RTCA defines as the removal of restrictions on users' flight paths and speeds. RTCA, Inc., a private, not-for-profit organization that addresses requirements and technical concepts for aviation, will serve as the primary organization for FFP1 Stakeholders to review and evaluate the operational impacts of the FFP1 Core Capabilities. FFP1 is derived from RTCA recommendations for achieving Free Flight objectives. Specifically, FFP1 provides near-term Air Traffic Management (ATM) capabilities that can deliver early benefits to National Airspace System (NAS) users and service providers, leveraging proven technologies with needed procedural enhancements and appropriate standards. The FFP1 core capabilities are (*Government/Industry Operational Concept for the Evolution of Free Flight, Addendum 1,* RTCA, August 1998):

- Conflict Probe (CP) as represented by the User Request Evaluation Tool (URET)
- Traffic Management Advisor (TMA), Single Center (SC)
- Passive Final Approach Spacing Tool (pFAST)
- Collaborative Decision Making (CDM) with Airline Operations Centers (AOCs):
  - NAS Status Information (NASSI)
  - Enhanced Ground Delay Program (GDP-E)
  - Collaborative Routing (CR)
- Surface Movement Advisor (SMA)

The table on the following page describes FFP1 capabilities and their expected benefits.

Table 1-1. Description of FFP1 Capabilities from Addendum 1

Program/System	Capability	Domain Used	Used by Whom	Benefits
URET	- Aircraft-to-Aircraft CP - Aircraft-to-Airspace CP	Enroute	D-Controller	- Controller decision aid - Reduce altitude and speed restrictions
Passive FAST	- Runway Assignment - Arrival Sequencing	Terminal	Controller, TMC	- Efficient use of runway capacity - Improved safety through better situational awareness
TMA	- Arrival scheduling - Arrival sequencing	Enroute, Terminal	TMC, R-Controller	- Load balancing between feeder fixes - Optimizes runway usage
SMA	- Aircraft Surveillance - Information to AOC ramp	Terminal	AOC, (Ramp Tower)	- Optimizes ground and ramp resources -More efficient planning
CDM and NAS Information	- NAS Status Information - Enhanced GDP - Collaborative Routing	Pre-Flight Planning, Enroute	ATCSCC, TMC, AOC, Controller	- Better planning by all NAS participants - User control of departure times - Collaboratively planned solutions when excess demand in system

The Free Flight Phase One Program Office (FFP1 PO) will deploy these capabilities which are currently in development or in limited operational use to selected sites. The FFP1 Metrics Team is comprised of experts from FAA and organizations with unique operational and analytical experience. These include the Center for Advanced Aviation System Development (CAASD), National Center of Excellence for Aviation Operations Research (NEXTOR), TRW, and other support contractors with air traffic experience. The Metrics Team will monitor changes in NAS performance associated with FFP1 capabilities using measures that represent the operational outcomes desired by system users. The operational outcomes used for FFP1 evaluation are adapted from other FAA publications, namely the FAA Strategic Plan and the ATS Performance Plan. These user-based outcomes are represented by the following performance categories:

- Safety
- User Access
- Delay/efficiency
- Predictability
- Flexibility
- System Productivity

1.2 Purpose and Scope of Plan and Evaluation

The FFP1 PO will collaborate with Stakeholders to evaluate the operational impacts of the FFP1 capabilities. This *Metrics Plan* explains the proposed FFP1/Stakeholders approach to evaluating the impacts/benefits of FFP1 to airspace system users and service providers. Stakeholders, for FFP1 evaluation purposes, will be represented primarily by the RTCA who "sponsored" the original Free Flight concept. The *Metrics Plan* describes FFP1 capabilities and expected impacts, performance metrics, data sources, evaluation schedules, and the analysis techniques to be used for comparing baseline versus in-use data. It also identifies issues associated with the evaluation and interpretation of data and results for each tool.

The overarching purpose of the *Metrics Plan* is to enable Stakeholders and the FFP1 PO to have a common framework for identifying and understanding FFP1 operational impacts. Core Metrics have been previously established by a combined RTCA/FFP1 team. These metrics were built upon the FAA operational goals and the expected FFP1 operational impacts. They are the basis for this *Metrics Plan*. As more experience with FFP1 capabilities is gained, the performance metrics will evolve. The FFP1/Stakeholders collaborative relationship will continue throughout the evaluation. Stakeholders' data inputs and "interpretation/validation" of impacts are a vital link to operational impacts. The FFP1 PO must make data available for the Stakeholders' review. The FFP1 PO will provide Stakeholders with consistent information from reliable sources to assess operational performance. The information will also facilitate future decisions about system enhancements, site proliferation, and funding.

FFP1 performance metrics are quantitative measures of operational impacts related to the FFP1 capabilities. They are measures of changes in activity, including but not limited to: actual arrival rates, flying time and distance for flight segments, and productivity. The "value" of these activity changes may relate only to specific peak time periods. Evaluating these performance metrics provides the means to associate operational impacts with economic value. The evaluation process must equitably assess impacts on the user community and ensure that all appropriate Stakeholders are included in a structured review.

Note that performance metrics can differ from programmatic metrics. Programmatic metrics assess whether a program or tool attains its intended function: the maturity, risk, and functionality of the capability itself. An example of a programmatic metric might be the metric to assess whether pFAST is sequencing aircraft efficiently (that is, measuring whether the pFAST algorithm to perform this function is adapted properly). The results of these programmatic metrics are important to the operational impact evaluation, to show causality of the change in NAS performance. The Metrics Team will work closely with the individual FFP1 Product Teams as they conduct their analyses, in order to associate tool performance with operational impacts.

All of the performance measurement and analysis in this *Metrics Plan* is in consonance with the guidance and direction of the *FFP1 Program Master Plan: Update* (PMP). The

PMP provides both the baseline and an overarching program management approach for the deployment of FFP1 capabilities in the NAS. The metrics are primarily based upon the *Government/Industry Operational Concept for the Evolution of Free Flight, Addendum 1: Free Flight Phase 1* (hereafter referred to as *Addendum 1*). Further validation of the metrics will occur as the data collection begins. Each of the metrics will remain flexible, and they will be refined as a direct result of feedback from users. It is expected that additional metrics will be incorporated into future versions of this plan, especially as a result of observing future benefits from tools deployed towards the very end of the FFP1 timeframe.

The Metrics Team acknowledges that it will be unlikely to get a unanimous agreement on the interpretation of each of the metrics for each FFP1 program. Given that the NAS is a very dynamic system, a change in the value of any of the metrics (either up or down) may be hard to definitively associate with a program. The Metrics Team will examine a large volume of data, including "state of the NAS" data (e.g., weather, block times, day of the week, level of demand, etc.) to aid in the analysis of FFP1 benefits.

The primary objective of the FFP1 operational evaluation is to support future decisions regarding NAS-wide implementation or future enhancements of FFP1 capabilities. Additionally, the evaluation results will provide inputs to the FFP1 PO on whether the expected performance impacts are being achieved. An auxiliary purpose of the operational evaluation is to gain new insights regarding operational impacts that may not have been anticipated. Operational evaluation activities range from capturing the performance of individual FFP1 Core Capabilities to understanding the site-level interactions of multiple capabilities. While not the primary focus of this evaluation, the Metrics Team will consider operational impacts on parts of the NAS beyond the FFP1 sites. The results of the operational evaluation will be shared with all FFP1 Stakeholders.

1.3 Data Requirements

Evaluation of the FFP1 Core Capabilities will depend upon the availability of operational performance data from a variety of sources. The FAA has access to a number of existing data sources, including Enhanced Traffic Management System (ETMS) data, Airline Service Quality Performance (ASQP) data, Consolidated Operations and Delay Analysis System (CODAS), and Automated Radar Terminal System (ARTS) track data. Air carrier data on fuel usage, causes and magnitudes of delays, numbers and causes of diversions and cancellations, numbers of refiled flight plans, and other performance factors will require inputs from Stakeholders.

Data to recreate the conditions surrounding the measurement and collection of the performance data will be needed. Context data include weather information (airport surface observations, severe weather areas, etc.), airport configuration, airport meteorological conditions (IMC, VMC), demand, and other contributing factors. Where possible, data sources have been identified for these context data. Additionally, site-specific data will be incorporated as the Metrics Team begins the validation and demonstration process.

1.4 Reporting Requirements

The FFP1 PO will implement both formal and informal reporting mechanisms to share the results of operational evaluations with Stakeholders. Formal mechanisms will include quarterly reporting to the RTCA Free Flight Steering. This reporting will be coordinated with the RTCA Select Committee on Free Flight Implementation and the Stakeholder subgroups.

Informal reports will range from anecdotal descriptions of operational impacts to responses to special data requests from Stakeholders. Data may be shared informally with Stakeholders so that Stakeholder assistance in data interpretation can be obtained. The informal reports, while more responsive to the immediate needs of Stakeholders, will not be subject to the same level of scrutiny as the formal reports, and therefore may not represent conclusive operational assessment results. Rather, they will serve to characterize the observed performance trends of the FFP1 Core Capabilities and be used to provide insight for more rigorous data collection and assessment.

1.5 Organization

This *Metrics Plan* is structured as follows. Section 2 describes each of the FFP1 capabilities. The third section presents an overview of the evaluation that is intended for each of the FFP1 capabilities. It defines an overall evaluation schedule and introduces the FFP1 performance metrics. Section 4 presents the evaluation methodology to assess the operational impacts of each of the FFP1 capabilities using the stated performance metrics. The final section describes the roles and responsibilities of the key players in the FFP1 evaluation process. The RTCA Core Performance Metrics (December 1998) are included as Appendix A. Appendix B describes the changes to the performance metrics from previous draft versions of the Metrics Plan.

Section 2

FFP1 Program Overview

2.1 FFP1 Capabilities

2.1.1 User Request Evaluation Tool

The User Request Evaluation Tool (URET) Core Capability Limited Deployment (CCLD) is a decision support tool that assists air traffic control service providers in meeting the needs of airspace users. For FFP1, URET CCLD general capabilities will be aircraft-to-aircraft and aircraft-to-airspace conflict detection and trial planning of proposed solutions to ensure that they are conflict free. Capabilities will be used primarily by the D-Controller for strategic problem detection (defined generally as a 20-minute look-ahead period). These capabilities will allow the sector team to approve more user requests and impose fewer altitude and speed restrictions in the participating sectors. The basis for URET CCLD strategic planning capabilities is information on the aircraft's flight intent, including flight plan information, track data, forecasted winds and temperatures, aircraft performance characteristics, and facility adaptation. Using this information, the progress of an aircraft is continuously monitored, problems are detected, and controllers are notified of possible conflicts between the current flight and other aircraft and/or adapted airspace. In addition, when the pilot requests a new clearance, the controller can use URET CCLD to identify any possible conflicts.

2.1.2 Traffic Management Advisor – Single Center

The TMA Single Center component of the Center TRACON Automation System (CTAS) assists controllers in the enroute cruise and transition airspace. TMA provides ARTCC personnel with a means of optimizing the arrival throughput of airports. By optimizing throughput TMA helps to reduce delays in the extended terminal area (200nm from the arrival airport) with respect to the ARTCC boundary. In situations where the serviced TRACON is within 200nm of the edge of the ARTCC airspace, there will be a provision to obtain flight track information from the particular adjoining ARTCC for the purpose of knowing the nature of the arrival flow out to 200nm. This is a step towards TMA Multi-Center. Inputs to the TMA system include real-time radar track data (i.e., aircraft position in three dimensions), flight plan data, and local meteorological conditions. TMA's trajectory models use this information, updated every 12 seconds, to compute routes and optimal schedules to the meter fixes for all arriving aircraft which have filed IFR flight plans, with consideration given to separation, airspace, and airport constraints.

The TMA computer interface incorporates two primary displays. The Timeline Graphical User Interface (T-GUI) displays estimated time of arrival, CTAS delay-imposed scheduled time of arrival, per aircraft delay, and runway assignment for each track in the TMA area of regard. The Planview Graphical User Interface (P-GUI) displays a planview depiction of arriving aircraft.

2.1.3 Passive Final Approach Spacing Tool

The pFAST component of CTAS is used by controllers and air traffic managers to manage the flow of arrivals through terminal airspace. pFAST computes a relative sequence for each arrival aircraft for each runway at the particular airport. The system calculates a near-optimal runway assignment for each aircraft in such a way as to minimize overall delay and increase airport throughput, with consideration given to aircraft type, speed, and trajectory. Runway advisories are then displayed to the controller as a three character runway identification that is time-shared with the sequence number on a third line that is added to the data block of the ARTS display. The controller may manually override both the relative sequence number and the runway advisory displayed by pFAST. The system automatically adjusts the sequence number, once the aircraft is committed to final approach. The anticipated impact of pFAST is more efficient use of both arrival and departure runways during peak traffic periods. pFAST displays also enhance a controller's situational awareness, especially during periods of heavy terminal operations.

2.1.4 CDM: Enhanced Ground Delay Program

Collaborative Decision Making (CDM) was conceived out of the FAA's Airline Data Exchange (FADE) experiments that began in 1993. These experiments proved that having airlines send updated schedule information to the FAA could improve air traffic management decision making. CDM has evolved from these same principles in an effort to improve air traffic management through information exchange and data sharing.

The initial focus of CDM, known as Enhanced Ground Delay Program (GDP-E), started prototype operations at San Francisco (SFO) and Newark (EWR) airports in January 1998. Under GDP-E, participating airlines send operational schedules and changes to schedules to the Air Traffic Control Systems Command Center (ATCSCC) on a continuous basis. This schedule information includes, but is not limited to, flight delay information, cancellations, and newly created flights. The ATCSCC uses this information to better implement and manage ground delay programs (GDPs).

GDP-E provides a more accurate view of demand, and it enables airlines to watch over and participate in ATM actions which directly affect their operations. Providing for simplified substitutions, control by arrival times, and daily download of flight schedules improves decision making, thereby reducing delays, unused slots, and needless modifications to schedules.

2.1.5 CDM: NAS Status Information

The NAS Status Information (NASSI) function will provide a mechanism to share safety and efficiency data with NAS users.

A mature NASSI capability is a robust information-sharing system. In the FFP1 timeframe however, NASSI will demonstrate various information collection and distribution possibilities while providing limited though essential NAS information. NASSI will present data on airport and airspace status and conditions. While this information is already available within certain organizations, NASSI will extend existing information to a broader range of NAS participants.

2.1.6 CDM: Collaborative Routing

In addition to improving the execution of GDPs, CDM has been found to have application to other air traffic management problems, such as airspace congestion. Under FFP1, the Collaborative Routing (CR) function is intended to provide better information to airspace users and enhance decision making procedures. CR will alert NAS users to potential flow problems that might require rerouting or other flow management actions. This would allow users to prepare for possible effects on their operation in advance.

CR will be primarily an information sharing system during FFP1 and not reflect the extent of collaborative capabilities envisioned for later stages of programmed improvements. CR is actually a collection of technologies, which will enable the exchange of real-time traffic flow information and updates between AOCs, ATCSCC, and ARTCCs. This information exchange will ultimately facilitate the efficient coordination of aircraft routing strategies between these planning bodies.

2.1.7 Surface Movement Advisor

Surface operations are improved by the use of information sharing at some airports. This capability makes available to airport ramp control personnel aircraft identification and real-time position information for aircraft arrivals in the terminal area. The availability of real-time information via the Surface Movement Advisor (SMA) will result in less congestion, reduced taxi delays, and more efficient use of crew and gate services in the ramp area. Knowledge of delayed flights allows ramp personnel to divert those gate resources to other temporary duties. In addition, if a gate is not available, ramp personnel may notify ATC personnel by voice connections to hold early arrivals on the airport movement area. More efficient gate operations allow more efficient surface operations to be conducted.

SMA terminal radar data (ARTS Data Feed) is also being installed at the operations centers for some airlines. The SMA ARTS Data Feed in the airline's AOC provides better situational awareness of the arrivals within the terminal airspace. This has lead to early airline savings in reducing the number of diversions by enhancing their awareness of activity in TRACON airspace. Increased knowledge of aircraft arrival positions will lead to better decisions on whether or not to hold connecting flights at the ramp. Additional operational impacts are expected including reduced internal airline communications time and reduced frequency congestion.

2.2 FFP1 Implementation Sites

Table 2-3, from Addendum 1, lists the locations for implementation of FFP1 capabilities. Operational impact analysis will be conducted on a site-by-site basis using the performance metrics presented in this *Metrics Plan.*

Table 2-1. Sites for FFP1 Implementation

Centers	FREE FLIGHT PHASE 1 CAPABILITIES1, 2, 3				
	TMA (SC)	pFAST	URET CCLD	CDM with/AOCs	SMA
ZAU: Chicago	6	(ORD) 6	✓		✓(ORD) 7
ZFW: Fort Worth	✓	✓(DFW)			✓(DFW) 7
ZLA: Los Angeles	✓	✓(LAX)			
ZTL: Atlanta	✓	✓(ATL)	✓		✓(ATL)
ZID: Indianapolis			✓		
ZME: Memphis			✓		
ZDC: Washington			✓		
ZOB: Cleveland			✓		✓(DTW) 7
ZMP: Minneapolis	✓	✓(MSP)			
ZKC: Kansas City		✓(STL) 8	✓		
ZNY: New York	4	(JFK), (EWR), 5 (LGA), (PHL)			✓(EWR) 7 (PHL) 7, (TEB) 7
ZOA: Oakland	✓				
ZMA: Miami	✓				
ZDV: Denver	✓				
ATCSCC				✓	
AOCs				✓	

Legend: ✓ - Part of FFP1 | Feasibility Study | Planned Capability | (xxx) - Location

Notes:

1 Operational by 1998-2002

2 Assumes no impact on other programs

3 Risks need to be identified and a risk mitigation plan established and implemented in coordination with RTCA

4 In 1998, launch R&D program to develop tools to expedite arrivals in complex airspace - at PHL (outside funding scope of FFP1)

5 By June 1998, determine feasibility of implementing stand-alone pFAST without multi-center TMA, and implement accordingly (outside funding scope of FFP1)

6 Begin development after completion of current airspace review and design

7 Functionality different from SMA at Atlanta

8 Stand-alone pFAST without single-center TMA

Section 3

Evaluation Overview

3.1 Evaluation Schedule

The FFP1 evaluation schedule is shown in Figure 3-1 and Figure 3-2. The schedule, based on the FFP1 Site Delivery Schedule and FFP1 Management Schedule produced by the FFP1 PO Implementation Team, presents a site-by-site overview of the evaluation activities associated with each FFP1 core capability. Activities and milestones include prototype in-use data, baseline (pre-CCLD) data collection, Initial Daily Use (IDU), observation period (following IDU), Planned Capability Available (PCA), and in-use (CCLD) data collection.

Figure 3-1 presents the evaluation schedule for FFP1 prototypes. The FFP1 prototype sites are unique to FFP1 deployment and to the performance evaluation process. The Metrics Team will use these sites to validate the performance metrics presented in this Metrics Plan. Baseline data for these prototype sites may be difficult to obtain given that the prototypes have been operational for some time. There have been multiple studies conducted on the benefits of these prototypes by the FAA and industry. The Metrics Team will use these studies as well as make an attempt to measure a "before" and "after" on these sites and capabilities.

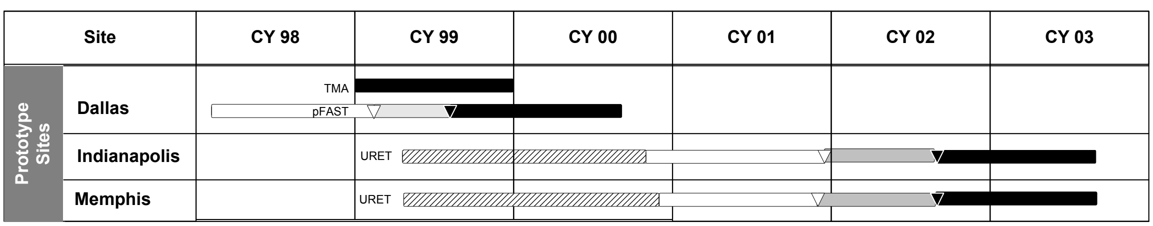


Figure 3-1. Integrated Evaluation Schedule (Prototypes)

Figure 3-2 presents the evaluation schedule for current CCLD sites and future CCLD deployments. The primary objective of the FFP1 evaluation is to capture a "before FFP1 capability" and "after FFP1 capability" representation of NAS performance using the metrics presented in this Metrics Plan. This is identified as a baseline data collection period and an in-use data collection period. The baseline data collection period is scheduled for one year prior to IDU. The Metrics Team will collect the data necessary to quantify the metrics presented in this Metrics Plan so as to obtain baseline performance at the specific site. For some of the FFP1 capabilities, this data collection will be done in conjunction with the adaptation data collection process. The sample size of one year data should provide the

Metrics Team with adequate data to normalize for the different factors that affect NAS performance.

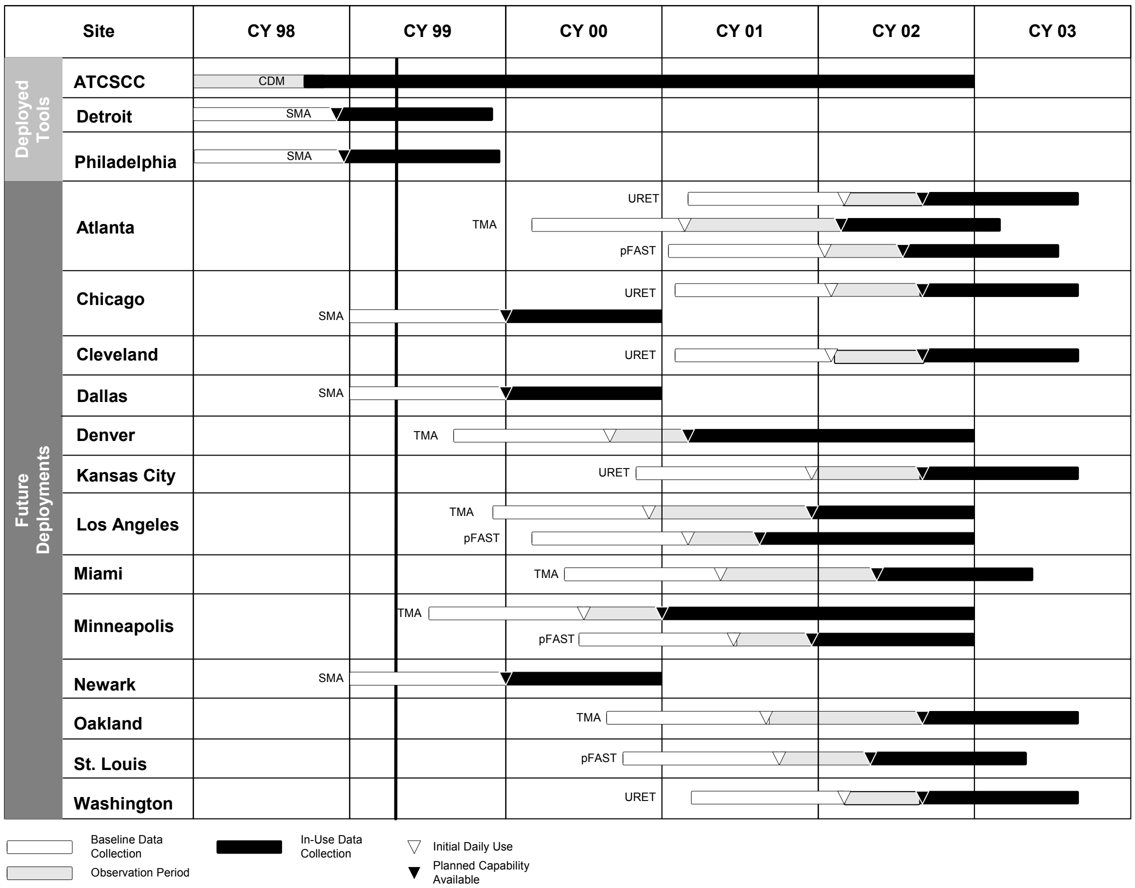


Figure 3-2. Integrated Evaluation Schedule (Current and Future Deployments)

The FFP1 PO has officially defined the milestones used for FFP1 implementation. Figure 3-3 explains these milestones. IDU signifies the hardware and software are installed and the initial cadre of operators are using the system to provide services to NAS users. Sd depicted in Figure 3-1 and Figure 3-2, the Metrics Team will continue to collect data through an “observation period” between IDU and PCA. During this time there may be no observable change, or possibly a degradation, in NAS performance as users are being trained and learning curves are taking effect. PCA signifies the planned cadre of operators are using the system on a regular basis to provide services to NAS users. It is at this point in time that the benefits analysis can be conducted. That is not to say that any benefits observed perform PCA will be minimized or disregarded.

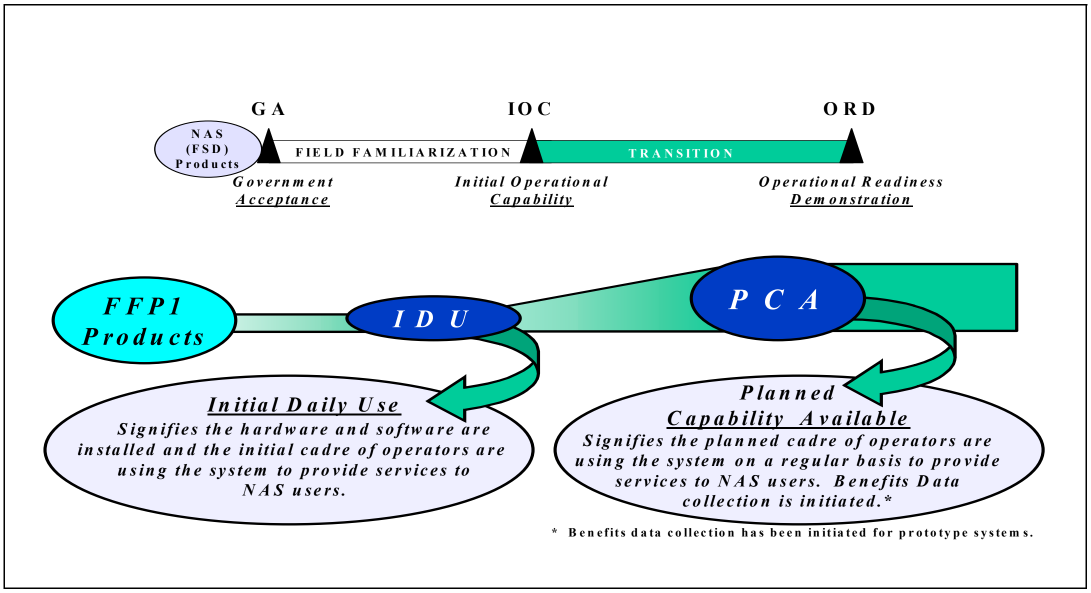


Figure 3-3. Stages of Deployment

The evaluation schedules presented in this Metrics Plan describe the evaluation process extending beyond the FFP1 timeframe (beyond CY02). For some FFP1 capabilities where deployment is late in the FFP1 timeframe the data collection, analysis, and valuation of the metrics data may extend beyond CY02. However, the current deployment schedule should lend itself to providing results for each capability at some subset of the FFP1 CCLD sites.

The specific roles and responsibilities for the tasks depicted in Figures 3-1 and 3-2 are described in section 5. In general, FFP1 metrics evaluation activities will require an unprecedented level of collaboration between the FAA and the user community. Roles and responsibilities will be more clearly defined as the metrics and required data are validated.

3.2 Categorizing Airspace User Objectives

In addition to maintaining metrics' traceability to operational impacts and capabilities, the FFP1 Metrics Team also recognizes the need to maintain their traceability to the FAA's well-established performance measurement framework, which is based on user-based operational outcomes. The FAA has identified four categories of system performance outcomes to measure capacity and efficiency: delay, access, predictability, and flexibility. For FFP1 performance measurement we have tailored these definitions in an attempt to reduce overlap. The Metrics Team has also identified areas which we believe can only be addressed (at a reasonable cost) through qualitative user inputs.

The underlying goal of FFP1 is to provide early benefits to NAS users using proven technologies. These benefits will be measured based on each capability's performance in achieving NAS user objectives. User objectives have been categorized as follows:

- Safety
- User access
- Delay/efficiency
- Predictability
- Flexibility
- System productivity

The specific definitions of these categories and the categories themselves have varied slightly when used by the FAA and the User community. To date, most definitions have not allowed for specific benefit generating events to fall into a single category. Capturing events in more than on category is not an issue unless value functions are created which could potentially double-count certain events.[1] For this reason the FFP1 Metrics Team has attempted to refine these metrics categories such that they are mutually exclusive. This will support the development of value functions that will support distinct results

Some objectives may differ from day to day. From a daily operations perspective "schedule integrity" seems to be the highest concern to airlines. Fuel efficiency for example is an overall airline goal but not at the expense of missing connections. Airlines would like a "predictable" system that could be viewed as providing consistent service in terms of flight times. Airlines also may desire the "flexibility" to be "predictable" where certain flights might be granted faster routes to make up for lost time at a gate. Here predictability is defined in terms of schedule integrity as opposed to consistent flight times.

The categories and definitions below are meant to clarify these terms for purposes of the FFP1 performance measurement task. We recognize that the definitions below do not completely solve the problem of overlap. The intent is to establish a method, which systematically groups or breaks down the components of each measure/metric, for valuation purposes and comparison.

The FFP1 Metrics Team recognizes that it may not always be possible to improve flexibility, predictability, access, and on-time performance simultaneously. There may be times when it will be necessary to trade off one against another. Moreover, the appropriate

[1] An airline wants the "flexibility" to take more "efficient" routes to reduce "delay"; or an airline wants to increase arrival rates to reduce delay during peak times; or "flexibility" to reduce flying time on a specific flight and "delay" other flights thus impacting "predictability".

tradeoffs will vary depending on user needs. Furthermore, the Metrics Team acknowledges that some aspects of operational performance may be difficult to measure. For example, while data can be easily obtained to measure delays directly, flexibility cannot be measured directly. Rather, indirect measures such as time or distance flown on desired routes must be used. Also, the establishment of measurement baselines to characterize "normal" operations will be challenging, as care will have to be taken to ensure that such baselines are unbiased with respect to seasonal effects, meteorological conditions, changes to the operational state of the NAS, and other factors.

3.2.1 Safety

Safety in the NAS will not be compromised as the FAA strives to improve flexibility, predictability, access, and on-time performance in the ATM system. The FAA's highest priority operational outcome is to improve system safety. The FAA has defined a set of safety standards that define spacing between multiple aircraft, aircraft and other physical structures, and aircraft and airspace. System safety, from an air traffic standpoint, is measured through the ability to maintain these standards. When aircraft violate these separation standards, an *operational error* occurs. Specifically, an operational error occurs when

- Less than the applicable separation minimum results between two or more aircraft, or between an aircraft and terrain or obstacles, or
- An aircraft lands or departs on a runway closed to aircraft operations after receiving air traffic authorization.

Similarly, when aircraft penetrate airspace that has not been pre-coordinated for that aircraft's use, *operational deviations* occur. An operational occurs when

- Less than the applicable separation minimum existed between an aircraft and adjacent airspace without prior approval, or
- An aircraft penetrated airspace that was delegated to another position of operation or another facility without prior coordination and approval, or
- An aircraft penetrated airspace that was delegated to another position of operation or another facility at an altitude or route contrary to the altitude or route requested and approved in direct coordination or as specified in a letter of agreement (LOA), pre-coordination, or internal procedure, or
- An aircraft, vehicle, equipment, or personnel encroached upon a landing area that was delegated to another position of operation without prior coordination and approval.

The intent is to track operational errors and deviations at all FFP1 sites and observe any changes in the existing rates. Where possible, baseline data will be segregated by conditions impacting the number of operational errors and deviations (e.g., weather, traffic density) and compared to similar conditions with the new FFP1 capabilities operating. Changes in the number of operational errors and deviations will be analyzed to determine any association with the FFP1 capabilities.

The FFP1 capabilities are intended to provide benefits to users while maintaining safe operating conditions. The expectation is that user benefits can be achieved while increasing the margin of safety. FFP1 capabilities provide controllers and airspace users with more information for increased situational awareness and improved decision making. FFP1 capabilities also provide an efficient and predictable sequence of aircraft from the enroute portion of flight to landing. These attributes of FFP1 support an increased margin of safety.

Observing the number of operational errors and deviations does not represent the efforts in FFP1 to address safety concerns. For each capability, safety issues are an integral part of the design and operation. From a FFP1 perspective, procedures, training, and reliability all address the safety impacts of each capability. FFP1 programs follow FAA Order 8040.4, Safety Risk Management, which directs specific steps to be taken to insure system safety. Safety is a major component of each operational test plan. The safety metrics are established only to assess whether the safety planning and analysis for each capability has been successful. The operational error metric and the operational deviation metric will also provide feedback to the FFP1 team regarding any further issues needing to be resolved.

3.2.2 User Access

User access can be defined as the ability of users to enter the ATC system and obtain services on demand. For FFP1 metrics purposes, access focuses on maximizing the use of existing runways for arrivals and departures. For CDM, access also includes system throughput related to improved information. Clearly, improved access to airspace and runways will have a direct relationship to delays. However, as demand increases, runway throughput may increase, while delays remain constant or potentially increase. This phenomenon is experienced with highways when a lane is added and drive time is initially reduced but increases with additional traffic (demand). For this reason, it is important to have specific measures for access (throughput) and delay.

3.2.3 Delay/Efficiency

Within the FAA and industry, delay has been defined as:

- Amount of time beyond expectations that it takes to complete a flight or individual flight segment.
- Time beyond the scheduled arrival time (flying public).
- Additional time above "optimal" or unimpeded time.

Each of these definitions has merit. For FFP1 purposes we will incorporate each of these interpretations since each has a unique impact on the NAS user's value function.

Definitions of "efficiency" have centered around fuel efficiency for a given flight or reductions in flight times. Efficiency has often incorporated all reductions in delay. From and FFP1 perspective, both have unique components which are valued separately as well as a common component. All will be captured under the "delay/efficiency" category.

The FFP1 metrics associated with this category are those capturing changes in flight segment times, differences in actual versus scheduled times, and fuel efficiency. Where FFP1 capabilities produce improvements in flight segment times the value can be expressed in both a fuel and time component. The time component has value to both variable (ADOC) and fixed costs (gates, airplane, etc.)

When analyzed for the total flight, reductions in time beyond schedule has value impacts in terms of "propagation of delay" and other penalties like passenger "payoffs". Additionally, there are flights that show no change from the baseline in-flight segment time but may have flown an improved route (by staying at more optimum altitudes or speeds).

For FFP1 purposes, a desired impact is to reduce pre-FFP1 flight segment times. Flight segment times prior to FFP1 are the baseline for measuring FFP1 impacts. Since theoretical "optimum" flight segment times will be equal for the baseline and the post-FFP1 observations, we can define delay reduction as the difference between the baseline and post FFP1 data. For example, if the optimal en route flight time for a specific flight under specific conditions is 125 minutes and the pre-FFP1 time is 135 minutes delay would be calculated at 10 minutes. If a similar flight under similar conditions post-FFP1 had an enroute flight time of 131 minutes the delay would be reduced to 6 minutes. Since we are controlling operations for equivalent flights (and aircraft) under like conditions, the improvement associated with FFP1 would be 4 minutes (Of course to have statistical significance we would collect numerous data points for this scenario).

For the purpose of the FFP1 metrics plan, the DELAY metric category will capture both schedule delay and the difference in flight segment times pre- and post-FFP1. Both measures are important to users. Schedule delay, however, can be influenced by airline scheduling practices and may have no correlation to the time a flight takes to move through the system.

3.2.4 Predictability

Predictability measures the variation in the ATM system as experienced by the user. For FFP1, predictability focuses on the variance associated with flight segment times described above under delay. Commercial airlines may benefit as much from a reduction in the variance (or an improvement in the consistency) of flight/taxi times as they would from a reduction in the overall average. Predictability allows for improved scheduling and more efficient bank operations.

Predictability can also be applied to the variance around scheduled arrival times. In some of these cases, improved schedule predictability may be completely counter to reducing the variance on flight segment times. For instance, a flight may experience a departure delay for a maintenance action and the airline dispatcher would like the pilot to make up time where possible. The pilot in turn may ask for direct routings not normally requested in the flight plan. As a result, it is possible to have faster enroute times then on average. Conversely, an airline may want a flight slowed down to allow a gate to open or to keep the integrity of a bank. The point is, on a daily basis an airline may utilize tactics that are counter to what is perceived as optimal from a strategic standpoint.

To address the situations described above, the FFP1 definition of predictability will be limited to variance around flight segment times. This metric is important from a strategic planning perspective and of particular interest to both ATC managers and airline schedulers. To capture the tactical interest of bank integrity independent of planners, "flexibility" will be used as the metric to capture the airlines' ability to meet dynamic needs.

3.2.5 Flexibility

Flexibility would ultimately measure the ability of the ATC system to meet users' changing needs in their efforts to optimize daily operations. For example, commercial air carriers may prefer more total delay in exchange for the on-time arrival of a specific flight with numerous connections. For FFP1 purposes, we have focused flexibility metrics on capturing what the user would like to accomplish on an individual flight basis, which is not already captured in the above metrics. Additional examples of flexibility include varying objectives on flights depending on whether there was an early or late departure and requests for altitude changes for passenger comfort.

From a practical standpoint, it is extremely difficult to establish airline intent on an individual flight basis. In fact, within an airline the pilot and dispatcher may have different objectives. For this reason, FFP1's primary approach to establishing a measure of flexibility is to separate flights into two categories: those delayed upon departure and those departing on time. The supposition is that those aircraft departing behind schedule will desire to make up time enroute. Flexibility will be measured by any post-FFP1 change in an airline's ability to make up time (to keep schedule). Other measures of flexibility will be addressed by obtaining feedback from the airlines on their perceived change/improvements in service.

3.3 Operational Impacts and Performance Metrics

In late 1997, the FAA Administrator's NAS Modernization Task Force recognized the need to focus on a set of already field-tested programs with great potential to improve the performance of the Air Traffic Management (ATM) system. The Task Force recommended that these programs be combined and managed in a separate Program Office to focus on timely fielding of operational benefits. Given that focus on operational benefits, the resulting FFP1 Core Capabilities are benefits-driven, as is this *Metrics Plan*.

Prior to the development of this evaluation plan, the RTCA Free Flight Select Committee's joint government-industry Metrics Working Group developed a process to define the performance metrics. The Working Group first identified the expected operational impacts associated with each of the FFP1 core capabilities. These impacts were aligned with the FAA and industry user-based operational outcomes described in Section 1.1. These impacts described the potential benefits of each capability and led to the development of an initial set of FFP1 performance metrics (RTCA Core Performance Metrics, December 1998, see Appendix A). From this recommended set of performance metrics, the FFP1 PO continued this process of metrics development. The Metrics Team began to investigate the usability of these metrics as well as developing additional metrics. This Metrics Plan presents the current set of FFP1 performance metrics that will be used to assess the operational impact of the FFP1 capabilities.

The FFP1 performance metrics are intended to guide the collection and analysis of data to support the evaluation of operational impacts. Each metric is formulated to help quantify the expected impact of a FFP1 core capability and is therefore directly traceable to a FFP1 capability and its expected impacts. All performance metrics presented in this *Metrics Plan* are likely to be refined and revised as the Metric Team gains further knowledge of each capability's functionality and the impact the capability has on NAS users. Currently identified performance metrics may be found to be insufficient in providing insight into FFP1 operational impacts, while additional metrics may be defined from unforeseen impacts that the users are observing. These two cases will be captured in future versions of this *Metrics Plan*. The FFP1 operational impacts and performance metrics are summarized in Table 3-1.

Table 3-1. FFP1 Expected Operational Impacts and Performance Metrics

Capability	Expected Operational Impact	Metric Category	Performance Metric
All capabilities	Level of aviation safety in the NAS will not be compromised at any time and must be maintained at a level equal to or exceeding current standards	Safety	Change in operational errors while capability is in use* Change in operational deviations while capability is in use*
URET	Use of a conflict probe reduces and perhaps eliminates the need for procedural restrictions that have been implemented to aid controllers in separating aircraft; Elimination of many ATC route and altitude restrictions allowing aircraft to fly user-desired paths reducing user costs	Delay/Efficiency	Average enroute time and distance flown (on-time departures), Average enroute air distance flown (on-time departures), Average fuel usage, Percentage of time spent at or near desired altitude for city pairs, Number of restrictions eliminated, Aggregate degrees turned
		Predictability	Average planned versus actual enroute time and distance flown in Center
URET	Approval of more pilot or airline requests allowing for fuel efficient routes/descents reducing user costs; Trial planning and replanning can identify when a request is conflict free and can be granted.	Flexibility	Average enroute time and distance flown (late departures), Average enroute air distance flown (late departures)
URET	Improved situational awareness enables the early notification of separation violations; Trial planning generates conflict free paths which reduces the workload associated with resolving future conflicts.	System Productivity	Number of aircraft per sector per unit time, Monitor alert threshold
TMA	Efficient utilization of Center/TRACON airspace through the implementation of flow strategies, according to the TMC's preferred metering method, to achieve maximized airport throughput; By scheduling all appropriate up-stream fixes in Center airspace a smoother traffic flow and equal distribution of delay among aircraft within the rush is achieved	Delay/Efficiency	Mean flight time from 200nmi range ring to meter fix, Mean arrival delay, Mean fuel usage from 200nmi range ring to meter fix, Variability of fuel usage from 200nmi range ring to meter fix
		Predictability	Mean error in predicted meter fix arrival time, Variability in error of predicted meter fix arrival time, Variability of actual arrival rate, Mean difference between airport acceptance rate and actual arrival rate, Variability of time from 200 nmi range ring to meter fix
TMA	Efficient utilization of runways by applying the right amount of pressure on the TRACON ensures that the TRACON's scheduling constraints are met but not exceeded, i.e. increased airport acceptance rate	User Access	Mean actual arrival rate
TMA	Decrease in controller workload and better workload distribution	System Productivity	Mean actual arrival rate/throughput per sector or position

Capability	Expected Operational Impact	Metric Category	Performance Metric
pFAST	Reduction in air traffic delay through more efficient aircraft sequencing in the TRACON area	Delay/Efficiency	Mean flight time from meter fix to runway threshold, Mean fuel usage from meter fix to threshold, Variability of fuel usage from meter fix to threshold
pFAST	Increased runway utilization	Predictability	Mean difference between airport acceptance rate and actual arrival rate, Variability of flight time from meter fix to runway threshold
		User Access	Mean actual arrival rate for each runway, Mean actual arrival rate
pFAST	Decrease in controller workload and better workload distribution	System Productivity	Distribution and throughput of operations per runway/position
GDP-E	Increased information and user flexibility (such as substitution, control by time of arrival, etc.) enables airlines to cancel flights and adjust schedules for airports with GDPs	Delay/Efficiency	Mean flight time, Compression minutes saved
		Predictability	Integrated Predictive Error (IPE), Rate Control Index (RCI), EDCT compliance ratio, Number of GDPs cancelled near start, Number of GDP revisions
		Flexibility	Mean distance flown, Control Time of Arrival
GDP-E	Fewer and shorter ground delay programs since airlines will resolve some problems themselves, results in increased utilization of reduced capacity airports	User Access	Number of airport operations, Number of unused slots, Number of cancellations
CR	Improved knowledge of system status allows airline dispatchers to be preemptive with schedule changes and route planning to avoid severe weather areas, congestion areas, etc. - airlines have better control over their schedules	Delay/Efficiency	Average flying time, Standard deviation of predicted fuel usage and actual fuel usage
		Flexibility	Number of user preferred routes flown, Average flying distance
		User Access	Number of operations, Number of diversions, Number of aircraft using SUA
NASSI	Increased airline understanding of ATM's intentions and actions which results in a decrease in workload or time savings in negotiations between Centers and SCC	TBD	TBD

Capability	Expected Operational Impact	Metric Category	Performance Metric
SMA	Increased surface movement efficiency	Delay/Efficiency	Mean taxi-in time, Mean taxi-out time, Mean gate delay
		Predictability	Variability of taxi-in time, Variability of taxi-out time, Variability of gate delay, Gate reassignment rate
SMA	Enhanced terminal data leads to increased situational awareness in TRACON airspace, reduced internal airline communications time, reduced frequency congestion, and fewer misunderstood communications	User Access	Diversion rate

* Further study is necessary by FFP1 PO, Stakeholders, and the FAA to determine if the impact on safety is a direct result of and/or clearly attributable to the capability.

3.4 Evaluation Methodology

This section outlines the approach to be used to identify performance impacts attributable to FFP1 capabilities. Performance impacts are measured through changes in the set of FFP1 metrics described in the previous section.. This methodology includes identification and collection of data, and analytical techniques to determine the impact of FFP1 capabilities under multiple operational conditions.

3.4.1 Data Collection

A thorough analysis of the impact of the FFP1 program will require a broad range of data from a variety of sources. Examples of these data sources include facility log files, URET data log files, ARTS and Host data, archived weather data and various NASDAC (National Aviation Safety Data Analysis Center) files. Examples of the specific metrics that will be drawn from these sources include airport arrival rates, taxi-out times and fuel usage from meter fix to runway threshold, as well as scenario variables like ceiling/visibility and runway configuration. A detailed description of the data to be collected can be found in section 4 of this document, under each of the FFP1 capabilities. The collection schedule for each capability is given in Figure 3-1 and Figure 3-2, Integrated Evaluation Schedule by Site.

3.4.2 Data Reduction and Analysis

The data required to support the analysis of FFP1 impacts will be extensive, requiring sophisticated data storage and manipulation technologies. Once a database management system is established, statistical analyses will then be required to evaluate the relationship between the candidate metrics, FFP1 capabilities and other system variables like weather, traffic density, traffic mix and runway configuration. Ultimately, the challenge will be to identify performance impacts (through the metrics) highly correlated with the new capabilities and independent of other variables in the system.

Using the data described in section 3.4.1, statistical techniques including Analysis of Variance (ANOVA) and multiple regression analysis can identify conditions which are correlated with the values of each metric of interest. Included in these conditions are weather, runway configuration, and presence of the FFP1 capabilities. In other words, ANOVA and regression analysis can be used to estimate changes in the metrics resulting from various system variables as well as the FFP1 capabilities. For example, if the ANOVA analysis showed that taxi-out times (the metric) changed in a statistically significant way when SMA (the capability) is in use, it would support the argument that the use of SMA tended to impact (presumably reduce) taxi-out times. ANOVA might also show that taxi-out times are also correlated to other operational conditions like visibility.

Whereas ANOVA focuses on correlation, regression analysis can provide an estimate of the magnitude and the direction of the change in a given metric as a particular system variable changes. Regression analysis, for example, could be used to estimate the SMA's impact on taxi-out times under poor conditions. Initially, ANOVA and regression analysis will focus on baseline data to understand changes in metrics caused by operational conditions; once data comparing performance with and without FFP1 capabilities are available, the impact of the FFP1 capabilities can be evaluated as well.

These results from these analytical techniques can be useful in estimating the economic impact of each of the FFP1 capabilities, and facilitate benefit estimates for potential future sites. They can also help clarify the impact of tools under different weather conditions, increased traffic density, and airport configurations.

As with all statistical methods, ANOVA and regression analysis can provide evidence in support of a specific hypothesis (e.g., SMA reduces taxi-out times), but they can never provide "proof." Further, they are sensitive to the intelligent choice of potential metrics, especially when the metrics are not independent of one another (e.g., taxi-out and taxi-in times). As a result, these methodologies must be used with care by analysts who understand the complexities of the processes being evaluated.

Please note that these analytical techniques attempt to take full advantage of the data set. More obvious comparisons of like days (peak periods) measured before and after the implementation of FFP1 capabilities will also be used in the initial analyses. The statistical techniques described, however, provide a more robust use of the data available.

3.5 Economic Valuation

Economic Valuation is a means to interpret operational impacts captured by the performance metrics. Valuation uses a common scale - $ - allowing comparisons of the relative value (benefit) of each capability at different sites under varying operating conditions. Additionally, economic benefits tied to cost estimates provide insights into tradeoffs with other FAA capital investments. Valuation will support future benefit-cost analyses related to further deployment of FFP1 tools. Developing benefit estimates (for both FFP1 sites and future sites) requires significant participation by the user community. The FAA's established practice is to base benefit estimates on Aircraft Direct Operating Costs (ADOC) and Passenger Value of Time (PVT). Recent studies have shown that the value of improvements (or disruptions) in the NAS go beyond ADOC and PVT which associate benefits strictly with reduced time (delays). Additional benefits from modernization capabilities such as FFP1 range from improved utilization of users' fixed assets to reduced environmental impacts.

Economic valuation of the metrics must be considered to fully understand the utility of each capability. It is possible to have multiple metrics for an FFP1 capability depicting diverging results. For example, access could be increased while delays go up slightly. Again, in these instances we will work closely with users/Stakeholders to better understand the relative benefits. We may find that through slight changes in the use of an FFP1 capability a desired balance can be achieved.

While economic valuation leading to quantified benefits is a goal for the FFP1 metrics team, it is understood that many anecdotally observed impacts may not lend themselves to quantification. We may also find that we are not able to collect baseline conditions similar to conditions where the tools/capabilities are in use. In these instances it will be a qualitative judgement as to how a scenario would have unfolded prior to the FFP1 capability being in place. “Diversions” are an excellent example where particular FFP1 tools should assist in avoiding diversions but the number of total diversions are not reduced. Once these subjective judgements are made, however, we can often “value” the impact.

It is anticipated that the economic evaluation of FFP1 capabilities will evolve with experience. Where appropriate, the Metrics Team will quantify estimates and use ranges based on sensitivity analyses to express uncertainty.

3.6 Environmental Impacts

FFP1 is expected to increase the fuel efficiency of flights through direct/wind optimal routings and enabling more efficient altitudes. Aircraft emissions can be directly linked to fuel consumption. The Emissions Division of the FAA's Office of Environment and Energy (AEE) is responsible for the policy, regulatory, and technical aspects of aviation air emissions as they relate to engine emissions, local air quality, and global atmospheric effects. As such, the FAA is interested in the potential impact of any large-scale program which provides improved fuel efficiency and reduced emissions.

On a broader scale, the FAA is participating on the International Civil Aviation Organization (ICAO) Committee on Aviation Environmental Protection (CAEP). CAEP is charged with the development of international standards and recommends practices for measuring and controlling aircraft noise and engine emissions. CAEP has recently expanded its consideration to include operational measures that have the potential to reduce aviation emissions. This interest has prompted the development of formal emissions analysis resulting from the deployment of new and enhanced FAA capabilities.

The September 1998 report, "The Impact of National Airspace Systems (NAS) Modernization on Aircraft Emissions", prepared by the FAA included an estimate of potential emissions reductions associated with NAS modernization as well as a methodology for analyzing potential annual fuel savings and associated emissions reductions. The methodology used in the above report will be incorporated into future ICAO sponsored studies on emissions. The FFP1 Metrics Team will also adopt this methodology in assessing the impact of FFP1 capabilities on emissions and continue to interface with AEE and ICAO as appropriate.

Methodology Overview

Aircraft produce air pollutants as part of their normal combustion process. The rate that these pollutants are emitted depends upon several factors including the ambient air temperature and thrust of the engine. Therefore, the amount of pollutants that an aircraft discharges will depend on the phase and duration of flight. Among the air pollutants discharged from aircraft three are considered criteria air pollutants and are regulated by the Clean Air Act (1970) and Clean Air Act Amendments of 1977. These include nitrogen oxides (NOx), hydrocarbons (HC), and carbon monoxide (CO). Additional pollutants produced by aircraft engine combustion include carbon dioxide (CO2), and sulfur dioxide (SO2).

The Emissions Report estimated that 94 percent of potential fuel savings will occur in phases of flight above 3000 feet AGL with the remainder occurring below that level. This combined fuel savings translates to an annual reduction in emissions of over 209 million pounds of NOx, 211 million pounds of CO, and 59 million pounds of HC. Carbon dioxide and sulfur dioxide savings were not estimated. Since FFP1 provides capabilities for the en route phase of flight (i.e. URET, TMA, Collaborative Routing), it is likely that the potential benefits from these capabilities will reduce emissions of criteria air pollutants. Under the delay/efficiency metrics noted in this Metrics Plan, fuel savings resulting from the use of each FFP1 capability will be evaluated. These fuel savings will be calculated according to phases of flight and translated into emission reduction. The aggregate emission reduction will then be broken out by type of pollutant.

Section 4

Plan for Operational Evaluation of FFP1 Capabilities

4.1 User Request Evaluation Tool

4.1.1 Evaluation Overview

URET provides enroute air traffic controllers with an automated conflict prediction capability to evaluate flight trajectories of current or proposed flights by identifying potential aircraft-to-aircraft and aircraft-to-airspace conflicts. The capability supports the enroute center sector team in predicting and resolving potential conflicts involving aircraft that are flying in low, high, or super-high sectors. The capability supports aircraft-to-aircraft analysis as well as the assessment of trajectories to detect unauthorized flight into a special use airspace (SUA). The availability of this ground-based conflict prediction capability can potentially help the FAA relax some restrictions, thereby enabling NAS users to fly more advantageous trajectories.

Where implemented, URET CCLD will provide the following capabilities:

- Automatic Conflict Detection
- Trial Planning for assistance with conflict resolution and granting user requests
- Flight Data management capabilities
- Auto Coordination

URET systematically checks for conflicts between aircraft, and between aircraft and airspace. After detection, URET provides the controller (D-side) with a visual notification of the predicted conflict. Conflicts can be resolved with the help of URET's *Trial Planning* function. The controller creates candidate trajectory changes that are checked by the Trial Planning function to ensure that they are conflict-free. Trial Planning extends URET's basic conflict detection capability into a versatile strategic decision support tool. Using URET's strategic notification and trial planning capabilities, a controller has more lead time to assess traffic situations and identify appropriate conflict-free resolutions. The additional lead time may allow a controller to more accurately assess and confidently approve more pilot-requested flight plan amendments, knowing they are conflict-free. With the two-way Host interface connection, a URET Trial Plan can be automatically entered into the Host.

Aircraft performance and intent information from the current flight plan are inputs to URET's modeling and algorithmic logic. URET models an aircraft's trajectory, places conformance bounds on the flight path, and routinely checks the aircraft position. When aircraft exceed the established conformance bounds, URET re-conforms the aircraft and updates the trajectory, which is re-probed for possible conflicts against all traffic.

Before attempting to quantify URET CCLD benefits, one must understand that URET provides some benefits and enables others. URET provides benefits by 1) providing an earlier and more accurate warning of potential conflicts than the existing set of tools at the enroute sector and 2) providing the capability to build resolution maneuvers that are conflict free for up to 20 minutes. Both features reduce the actions a controller and pilot must take to resolve conflicts. URET enables benefits by facilitating a reduction in restrictions. The NAS imposes route and altitude restrictions to help controllers separate aircraft from one another. However, while restrictions may help controllers, they generally prevent users from operating their aircraft along efficient paths. Since URET provides automation assistance to help controllers predict conflicts, URET can enable the elimination of some restrictions in enroute airspace. Consequently, users can fly closer to the routes and altitudes they prefer. By approximating what the user prefers (e.g., for North American Route Program (NRP) flights, users file a flight plan with their desired route), benefits can be estimated by measuring how well the NAS allows users to do what they want to do.

The metrics described in the following paragraphs are candidates for evaluating enroute domain benefits during the FFP1 period. Some attempt to measure URET-provided benefits while others measure URET-enabled benefits (as differentiated above) by assessing how well the system allows users to do what they want to do.

4.1.2 Performance Metrics

Since URET is a controller tool, much of its functionality is invisible to the airline users. URET will provide controllers the ability to do their jobs more efficiently, which is expressed through output metrics, mostly those that quantify productivity categories. Controllers will be better able to respond to user preferences, the results of which are expected appear in outcome metrics. The outcomes most directly affected are described by delay/efficiency, predictability, flexibility, and productivity categories. Table 4-1 identifies the metrics and their relationship to FAA operational outcomes.

Table 4-1. URET Performance Metrics

Outcome Category	Metric
Safety	Change in operational errors while capability is in use
	Change in operational deviations while capability is in use
Delay/Efficiency	Average enroute time and distance flown (on-time departures)
	Average enroute air distance flown (on-time departures)
	Average fuel usage
	Percentage of time spent at or near desired altitude for city pairs
	Number of restrictions eliminated
	Aggregate degrees turned
Predictability	Planned versus actual enroute time and distance flown in center
Flexibility	Average enroute time and distance flown (late departures)
	Average enroute air distance flown (late departures)
System Productivity	Number of aircraft per sector per unit time
	Change in monitor alert threshold

4.1.2.1 Safety

With the deployment of URET, the level of aviation safety in the NAS is expected to be maintained at a level equal to or improved. Thus, two metrics, the change in operational errors and the change in operational deviations, will be used to analyze the possible operational affects of URET to NAS safety.

Operational Errors

The change in operational errors measures the ability to maintain standards for separation between aircraft and aircraft. An operational error occurs when an aircraft pair violates a separation minimum. Since URET provides an earlier and more accurate warning of potential conflicts than the existing set of tools at the enroute sector, it should contribute to a reduction in operational errors. Analysis of operations with the URET prototype currently fielded has shown URET to provide more warning in almost every case.

In quantifying this metric, the facility records needed to determine the number of operational errors as well as the improved warning time provided by URET may be difficult to obtain. Also, conflict alert is a last-minute tactical tool intended to prevent near mid-air collisions, whereas URET is a strategic tool intended to prevent operational errors. Therefore, the comparison is not like-to-like. Finally, the number of operational errors is small; it might be difficult to find any statistical significance in these numbers over the evaluation period.

Operational Deviations

The change in operational deviations measures the ability to maintain procedural separation between aircraft and airspace. An operational deviation occurs when an aircraft operates in an airspace not "owned" by the aircraft's controller (including un-coordinated entrance to another controller's sector.

In quantifying this metric, the facility records needed to determine the number of operational deviations may be difficult to obtain. In addition, operational deviations are not tracked by automation, but they are manually reported. It is possible that the number of operational deviations increases due to several other factors such as increased reporting of incidences. An increase may also indicate that the airspace structure needs to adjusted to accommodate the changing traffic patterns due to Free Flight (e.g., more directs filed; less unnecessary maneuvers, "gentler" maneuvers that span multiple sectors).

4.1.2.2 System Delay/Efficiency

Delay/efficiency metrics measure the amount of time, distance, or fuel beyond expectations that it takes to complete an operation. Both scheduled and actual flight times will be examined for any impact.

Average Enroute Time and Distance Flown

The time an aircraft spends in an airspace depends on the distance it must fly and its speed. With a conflict probe capability, the FAA may facilitate the removal of route restrictions and allow NAS users to fly more direct routes between city pairs. [Although the benefit is more directly due to the relaxed route restrictions, it has been postulated that URET's conflict probe may facilitate the removal of restrictions; thus this metric is included here.] The FAA can also reduce the distance an aircraft must fly for conflict resolutions by using URET trial planning to make efficient reroutes. In either case, the reduced distance can translate into a decrease in the time flown in URET sectors.

After deployment of URET CCLD to the seven FFP1 locations, flights operating between city pairs within this airspace would be the best candidates for calculating the average enroute time and distance flown and comparing these figures to analogous ones calculated before URET deployment. Due to the many maneuvers associated with climbs and descents, these measurements should be taken for the cruise portion of flight. The enroute time and distance can be calculated by subtracting the time when the aircraft is 40 nautical miles (nmi) from the departure airport from the time when the aircraft is 40 nmi from the arrival airport for each city pair. An average of the times and distances over a period can then be calculated by aircraft type.

Flight time and distance are also affected by aircraft speed and pilot preference. For example, the pilot of an aircraft that departs on time would probably operate the aircraft at a preplanned speed to minimize fuel consumption and arrive on time. In contrast, the pilot of an aircraft operating between the same city pair but departing later than scheduled may operate the aircraft at a speed that is greater than fuel-optimal to arrive on time and allow passengers to make their connections. To account for some of this variability, the enroute time for aircraft which depart later than scheduled will be separated from those which depart on time. The performance of ATC in responding to late departures is considered under the System Flexibility metrics category (Section 4.1.2.4).

There are several factors or issues that may affect the results of this metric. If users can routinely plan on flying along their preferred routes and reducing enroute times, they may alter flight schedules. However, once a schedule is set, users may still choose to modify aircraft speed to account for variations in departure times, wind, and weather, to enable their aircraft to arrive close to the scheduled time. Since the user may alter aircraft speed due to the situation and since speed affects the flight time, there might be considerable variation in time measurements, and it might not be clear whether a decrease in the mean was good or an increase was bad. Additional factors such as aircraft weight, winds, and weather can also impact times and distances. It may be hard to isolate all of these factors. For this metric as well as for others, it is conjectured that in the aggregate, many of the complicating factors are present in much the same quantities in the pre-URET measurements as in the post-URET measurements as, and so overall, the *difference* in the aggregate average time and distance traveled will still be meaningful.

Average Enroute Air Distance Flown

Air distance is a efficiency metric that measures how well the NAS allows users to do what they want to do. With the assistance of conflict probe, the FAA could allow users to file and fly the most desirable routes. So a reduction in air distance might be shown through statistical analysis to be attributable to URET.

Unlike a car, for which the distance traveled between two points is a simple ground distance, an aircraft travels a distance over the ground and through the air. Due to the effect of winds, these distances generally differ and are the same only in the cases when no headwind or tailwind component affects the aircraft. To minimize the fuel used in flight, NAS users want to take advantage of tailwinds or decrease the effect of headwinds. Planning routes to account for winds can routinely result in ground distances that are larger than great circle distances because the best wind may be located some distance away from the great circle route (i.e., the shortest distance between two points on a sphere).

The FFP1 capabilities that allow the FAA to reduce route restrictions will permit users to plan their flights along the routes they desire. The motivation behind a user's desire may not be clear, given the demands of the situation (such as speeding up or slowing down to meet a schedule). However, an aircraft operator will normally want to meet those demands in a fuel-efficient manner by taking advantage of tailwinds and mitigating the adverse effect of headwinds.

For a particular weight, speed, and altitude, an aircraft can fly a certain number of miles using a given quantity of fuel. This consumption rate may be expressed as miles per 1000 pounds of fuel. Since an aircraft operates in a dynamic air mass, the distance flown through the air mass or "air distance" is expressed as nautical air miles (NAM). Air distance is the product of the time it takes to fly a particular leg of a flight and the true (actual) airspeed flown through the air mass.

If users can reduce the number of nautical air miles flown they can save fuel. Therefore, calculating the air miles per flight between city pairs both before and after FFP1 capabilities are in place would be one way to measure how well the NAS allows users to do what they want to do. In the event that actual detailed fuel burn information is unavailable for all flights, a reduction in air distance should equate to a reduction in fuel. The following example illustrates this concept.

In this example, an aircraft flying at 400 knots true airspeed (KTAS) at FL 330 can fly 50 NAM for every 1000 lbs. of fuel expended. (For this simplistic example, a static fuel burn rate is used. In actuality, fuel burn continuously decreases at a given speed and altitude.) If this aircraft does not experience a headwind or tailwind between two points 1000 ground miles (nmi) apart, its ground speed will be 400 knots. The absence of wind effectively equates the aircraft's true airspeed and ground speed. Further:

Flight time	2.5 hours = 1000 nmi / 400 knots
Air distance	1000 NAM = 400 KTAS × 2.5 hours
Fuel usage	20,000 lbs = 1000 NAM × 1000 lbs/50 NAM

If the same aircraft flew the same route with a 100-knot tailwind, the aircraft's ground speed would equal 500 knots (400 KTAS + 100 knot tail wind). Consequently, flight time, air distance, and fuel usage all would be smaller:

Flight time	2.0 hours = 1000 nmi / 500 knots
Air distance	800 NAM = 400 KTAS × 2.0 hours
Fuel usage	16,000 lbs = 800 NAM × 1000 lbs/50 NAM

In a more practical example, assume the aircraft noted in the previous example is flying a great circle route. Additional assumptions include:

1) The ground distance between top-of-climb and top-of-descent is 1000 nmi.

2) There is a 100-knot tailwind north of this direct route. (For simplicity, the wind advantage applies to the entire 1200 nmi trip.)

3) To take advantage of the winds, the aircraft has to fly a ground distance of 1200 nmi.

If the longer ground distance route were chosen, the aircraft's ground speed would be 500 knots and it would take 2.4 hours to fly the 1200 nmi. However, the air distance would be 960 nmi and the aircraft would use 19,200 lbs. of fuel. Therefore, flying a ground distance that is 200 nmi longer would yield a fuel savings of 800 lbs.:

Flight time	2.4 hours = 1200 nmi / 500 knots
Air distance	960 NAM = 400 KTAS $\times$ 2.4 hours
Fuel usage	19,200 lbs = 960 NAM $\times$ 1000 lbs/50 NAM

Air distance will be calculated for flights between city pairs in URET airspace (e.g., Chicago to Atlanta). Ideally, due to the structure imposed during the departure and arrival phase of flight, this calculation should begin at top of climb and terminate at top of descent. For consistency and ease of analysis, the calculation of this metric will include only flights outside a 40 nmi ring around the departure and arrival airports.

The calculation of air distance will require data on the time it takes to fly various legs of a flight and the true airspeed on each of those legs. Further, calculating time requires either track or ETMS data to determine how long it took an aircraft to fly between a series of points. True airspeed calculations will require information on an aircraft's ground speed, track, and winds at the aircraft's altitude. A comparison of the aircraft trajectory and winds aloft is needed to calculate the headwind or tailwind component affecting the aircraft's forward progress. This factor is combined with the ground speed to calculate the aircraft's true airspeed. The true airspeed is then multiplied by the time in flight on each leg to arrive at the air distance traveled for that leg.

There are several factors or issues that may affect the results of this metric. The air distance flown between a given city pair varies due to enroute winds, which also vary on a daily basis. Even for a fixed city pair, a user-beneficial air distance flown on one day may be much greater than a user beneficial air distance on another. Therefore, using air distance to assess benefits requires a comparison of like "wind days". Again, it is conjectured that in the aggregate, the difference (between pre-URET and post-URET) in the aggregate average air distance traveled will still be meaningful.

It is assumed that the aircraft would file and fly wind-optimal routes if allowed. However, it is not common practice today, and both airlines and controllers need to change today's practices to fly these wind-optimal routes.

Average Fuel Usage

Aircraft fuel usage is one of the bottom line measures for NAS users. Capabilities (such as URET) that enable the elimination of restrictions should contribute to a user's ability to reduce fuel consumption by permitting more operations along wind-optimal routes and at optimum altitudes and airspeeds with minimal delay. The Metrics Team will collaborate with the Stakeholders to identify candidate routes or city pairs for use in the evaluation. During FFP1 evaluations, the Metrics Team will rely on the airlines to provide fuel use information for flights operating between specific city pairs.

In quantifying this metric, there may be some reluctance on the part of the airlines to release certain types of data. It may be difficult to obtain the fuel burn for only the enroute portion of flight to determine fuel savings due to URET.

Percentage of Time Spent at or Near Desired Altitude for City Pairs

This metric identifies the extent to which the system accommodates user altitude preferences. This metric assumes that users file their desired altitudes in the flight plan. Rather than just identifying the aggregate number of flights that attain the requested altitude, the metrics quality could be improved by also identifying the percentage of time per flight at the desired altitude. It has been postulated that URET will increase the time spend at higher altitudes for the following reasons:

- The altitude restrictions needed to separate flows of traffic may be reduced with the URET conflict probe.
- URET trial planning may provide the controller with a lateral maneuver when otherwise the controller may have used altitude separation (e.g., descending the arriving aircraft early to avoid other traffic).
- The URET conflict probe may prevent unnecessary maneuvering for aircraft which will approach each other but with a large separation (e.g., 15 miles). For example, without URET, departing aircraft may be given interim altitudes to pass below crossing traffic when lateral separation would have been large.

To calculate the percentage of time spent at desired altitude requires information about an aircraft's desired altitude; the altitude obtained from the filed flight plan will be used. A formal definition of this metric is as follows. First, identify a "cruising flight" by (1) constructing the empirical distribution of reported flight levels in the track; (2) find the mode "M" (the most common value) of this distribution; (3) compute the number of reported altitudes that are within 300 feet of M [the "300" is a parameter]; (4) if this number is more than one-half the total number of reported altitudes, then the flight is a "cruising flight". Next, for "cruising flights" only, calculate the number of altitude reports within 300 feet of the target cruise altitude given in the flight plan [which may or may not be the same as the mode "M" described above], divide by the total number of altitude reports, and express the result as a percentage. If implemented on a center-wide basis, this metric will be applied only to that position of a flight that is above a specified "floor" altitude. If implemented on a NAS-wide basis, this metric will use the aircraft's intersection with a 40-mile ring outside departure airport as a starting point and the intersection with a 40-mile ring outside the arrival airport as an end. A comparison that shows the percentage of the cruise portion of flight where the aircraft achieved its filed altitude can be made.

There are several factors or issues that may affect the results of this metric. To measure how well the NAS accommodates user-desired altitude requires knowing that altitude. It is assumed that users calculate their desired altitude based on aircraft performance and environmental conditions such as winds and perhaps turbulence, and the users file this altitude or a set of desired altitudes in an aircraft flight plan. However, in many cases pilots request changes to flight planned altitudes primarily due to turbulence. Therefore, comparing planned to actual altitudes may not reflect what the user wanted. To identify the altitude the user really wanted would require sector observation and/or the analysis of voice tapes, a time-consuming and resource intensive process.

Number of Restrictions Eliminated

To assure aircraft separation, controllers must predict the future location of aircraft. Route and altitude restrictions are in place to help controllers with this task. Unfortunately, the restrictions that help controllers generally prevent NAS users from operating along efficient paths. It would be inappropriate to strive to eliminate all restrictions. Restrictions allow orderly flight between heavy traffic flows. However, it is believed that some restrictions are too penalizing, that some restrictions applied too often (e.g., miles-in-trail restriction on Atlanta ARTCC traffic to Chicago O'Hare all day long, every day), or that restrictions are applied to remove some of the complexity of the traffic flows in certain sectors. The final type of restriction (to reduce airspace complexity) is where URET could provide benefits.

The URET CCLD provides automation to help controllers predict and develop resolutions to potential aircraft and airspace conflicts. The successful deployment and regular use of this tool may enable the removal of certain restrictions. If this expected impact does occur, it is likely that restrictions would be relaxed on a case by case basis rather than en masse. However, since restrictions have been in place for many years, their removal may meet with some resistance.

Large numbers of removed restrictions are not expected during the FFP1 period. As a result, efforts will focus on recording and analyzing all relaxed restrictions for individual aircraft during the study period. Another possible direction would be to examine those flights which departed after schedule time since pilots of late flights frequently request expeditious routings.

Identifying the number of documented restrictions eliminated requires the review of facility records to include standard operating procedures (SOPs) and letters of agreement (LOAs). Once identified, eliminated restrictions should be assessed for impact on the NAS. Studies will be conducted at the URET prototype sites to determine the impact of the removal of certain restrictions.

In quantifying this metric, it may be very difficult to attribute the elimination of some restrictions specifically and exclusively to URET. For example, the FAA has already eliminated many route restrictions as part of NRP without the advantage of a conflict probe capability. However, it may be easier to attribute an elimination of “individual case” altitude restrictions to URET. Unfortunately, identifying such cases is more cumbersome than identifying LOAs or SOPs altitude restrictions because there may be no record other than a review of track data or voice recordings.

Aggregate Degrees Turned

In the pre-FFP1 environment, air traffic controllers tactically manage traffic, resulting in ATC-directed maneuvers to separate and sequence aircraft. Due to human perception limitations, a maneuver to avoid a conflict may result in a subsequent maneuver after a relatively short time. ATC-directed maneuvers to avoid conflicts affect controller and pilot workload and increase the distance an aircraft must fly. Since URET CCLD provides a trial planning capability that will extend a controller's ability to detect conflicts and provide clearance for problem-free maneuvers, both the total number and the total overall magnitudes of ATC-directed maneuvers theoretically should decrease

Aggregate degrees turned is a proxy for measuring ATC-directed maneuvers because the latter requires resource-intensive sector observation and/or the analysis of voice recordings. The metric quantifies the sum of absolute heading changes during a flight (i.e., a 30 degree turn right does not balance off a 30 degree turn left) in a center’s enroute airspace. For a given flight, it is normalized by dividing by the total nmi flown by that flight in that center, and so should be thought of as aggregate degrees turned per nmi of flight.

The rationale behind this metric is that an aircraft operating along a user preferred route would not need to make many heading alterations during flight if permitted to proceed uninterrupted. If there are a number of ATC-directed maneuvers, the sum of absolute heading changes would be greater than in situations where the flight was permitted to proceed closer to what the user had planned. Note, however, that ATC-directed altitude changes will not be tallied in this metric.

Since the sum of angles would create a negative bias against long routes, it is appropriate to normalize the angle total by the distance traveled so that the unit actually measured is degrees per aircraft per nautical mile. It is also appropriate to consider that URET operates in an enroute environment and that measuring angles turned at low altitudes would be complicated by holding patterns, terminal routing requirements, etc. Consequently, the application of the metric will be restricted to those parts of an aircraft's route which are at higher altitudes. Currently, this cutoff is FL 180.

The idea was to get one measure of the extent (over time) to which aircraft are being maneuvered within a center. Despite the presence of deliberate turns and even "good" turns that may originate in the given center as a direct to a point in another center, it is conjectured (as before) that in the aggregate, complicating factors are present in the pre-URET measurements in much the same quantities as they are in the post-URET measurements, and so overall, the difference in the aggregate number of degrees turned (per nautical mile) will still be meaningful.

Algorithmic Details of Aggregate Degrees Turned

The statistics derived from the track data include *dx/dt* and *dy/dt*, which are produced by the NAS tracker and are used to calculate the bearing as $tan^{-1}(dy/dx)$. Track data are generated in 12 second intervals and every N track points the bearing change is calculated by subtracting bearings: $b_{\Delta i} = b_i - b_{i-N}$, where b_i is the bearing at track point i. Currently the value of N is 6, which means that bearing changes are calculated over 1-minute intervals. (If N is set too small the process can become more subject to noise, and if N is set too large the bearing change can miss parts of turns.)

The three thresholds used to control the process are:

1) To distinguish the beginning of a "true" turn from small fluctuations in bearing, $b_{\Delta i}$ is required to exceed t_{min}, where t_{min} is a minimum bearing change over N points (currently set to 4.5 deg). Each succeeding segment with bearing change greater than t_{min} will result in the signed value of that bearing change being added to the total for the turn. A turn is completed when the bearing change $b_{\Delta i}$ falls below t_{min} for two successive segments.

2) At the completion of a turn, the resulting absolute value of the signed sum of the bearing changes has to exceed t_{accept} (currently set to 5 deg) in order for that turn to be considered large enough to not be considered a small fluctuation.

3) Finally, since small corrections at the end of a turn are common, distinguishing these corrections from reversals that signify the beginning of a new turn is important. When $b_{\Delta i}$ is signed differently from the preceding $b_{\Delta i}$, and when both the current signed sum and the new $b_{\Delta i}$ exceed $t_{reverse}$ (currently set to 15 deg), the ongoing turn is considered to be complete and a new turn begun.

This metric is a crude, comparative gauging of a change in ATC-directed maneuvers. It has been postulated that when the aggregate degrees turned (per nautical mile) increases across many flights in a scenario, the ATC-directed maneuvers have increased. However, this metric does not distinguish between, say, a 30 degree turn after which an aircraft returns to its original route after 10 nmi, and a 30 degree turn after which it returns after 50 nmi. Consequently, it is not possible to relate aggregate degrees turned to a dollar figure. Its value lies in its ability to gauge, on a comparative basis, the extent to which ATC-directed maneuvers are reduced.

Since the aggregate degrees turned metric employs the use of thresholds, it is subject to the classic problems of establishing a threshold. If thresholds are too high, areas of interest may be missed. If thresholds are too low, calculations may be overwhelmed by noise.

4.1.2.3 Predictability

Predictability metrics measure the variation in the ATM system as experienced by the user. T he metrics related to predictability are planned verses actual distance flown in center and planned verses actual time flown in center.

Planned versus Actual Enroute Time and Distance Flown

This metric is an alternative way to frame the average enroute time and distance flown by calculating the difference between the time and distance an aircraft planned to be in an airspace, and the actual time and distance flown in that airspace. This definition accounts for reductions in number and impact of ATC-directed maneuvers.

There are several factors or issues that may affect the results of this metric. Since the collection method for this metric is to collect data over a volume of airspace (e.g., an ARTCC), misinterpretations of raw data will occur. For example, if an aircraft is given a large lateral maneuver that causes it to leave the center early, the raw data would show a reduction in the actual versus predicted time and distance; and this event would be incorrectly interpreted as a benefit. On the other hand, as before, it is conjectured that in the aggregate, the difference (between pre-URET and post-URET) in the aggregate average air distance traveled will still be meaningful. But additional filtering of the raw data might be appropriate. The additional filtering could include only analyzing flights which met certain criteria (e.g., actual flown distance was within 20% of predicted distance, minimum distance in center needs to be greater than x nmi, aircraft must exit at its predicted exit sector or one of the sectors adjoining the exit sector, etc.).

4.1.2.4 Flexibility

Flexibility metrics measure the ability of the system to meet users' changing needs in their efforts to optimize daily operations.

Average Enroute Time and Distance Flown, and Average Enroute Air Distance Flown (late departures)

These metrics are similar to that described under Section 4.1.2.2 with the exception that they consider those aircraft which depart later than scheduled and thus may be attempting to make up time.

There are several factors or issues that may affect the results of this metric. In order to determine which aircraft departed later than scheduled, a variety of data sources contain the needed information (e.g., ASQP, OAG, ETMS, SAR). It may be difficult to consistently correlate scheduled departure time for a flight with an actual departure message.

4.1.2.5 System Productivity

Productivity metrics measure the rate of airspace operations that controllers, airspace, or airports can safely manage per period of time. One possible unit that may be considered is the number of sector operations per year. Of course, since the characteristics of sectors vary across the NAS, this unit would be sector specific.

Change in Monitor Alert Threshold

The Monitor Alert Threshold (M/A) compares demand on components of the NAS to the capacity of that component. M/A provides the capability to project traffic demand for all airports, sectors, and fixes of interest in the continental United States. In addition, it automatically generates alerts when the projected demand exceeds capacity alert thresholds. Alerts are provided in visual and aural form. Predictions are in 15 minute increments for up to 4 hours into the future.

For M/A, sector capacity is based on the number of aircraft that experience has shown that a controller can routinely manage. When comparing sectors, the M/A value can vary due to the sector geometry as well as the type of traffic traversing the sector. For example, the M/A threshold for super high sectors where most traffic maintains a constant altitude may be proportionally higher than that for a transition sector where there are aircraft in climbs and descents. The reason is the dynamics of the transition sector requires greater controller concentration to predict and resolve possible conflicts. URET CCLD should help controllers manage this type of traffic. This increased manageability in-turn could facilitate an increase in the M/A threshold.

Another use of M/A information is to determine how often M/A thresholds are exceeded. It has been postulated that URET, as part of the FFP1 suite, may smooth out traffic; thus reducing the volume of M/A events.

There are several factors or issues that could affect the result of this metric. As controllers become comfortable with URET's strategic conflict detection and trial planning capability there may be some changes in the monitor alert thresholds. However, changes may require considerable time unless there is some impetus to make the change. It is possible that the M/A thresholds might be unchanged throughout the data-collection timeframe; unchanged values will be considered a neutral result (neither benefit nor penalty for URET).

Number of Aircraft Per Sector Per Unit Time

The number of aircraft per sector per unit time could be used as an indicator of improved efficiency in delivery of services. In the aggregate, this metric could indicate if the center is handling increasing amounts of traffic.

The number of aircraft in a sector can be calculated by identifying aircraft for which:

1) track control is assigned to the sector,
2) voice communications is assigned to the sector, or
3) the aircraft is within the physical boundaries of the sector.

The first option uses on the handoff of track control to determine when an aircraft is in a sector. Getting the actual boundary crossing time of the sector requires track data, and getting the time of transfer of voice communications requires voice tape analysis. The track control method is the most automated way of calculating the time an aircraft is in a sector.

URET is intended to positively impact traffic on a scale larger than that in a single sector or small group of sectors. Restricting this metric to the sector level might be too small an area for the purposes of examining the real impact of URET.

4.1.3 Evaluation Schedule

The URET evaluation schedule is illustrated in Figure 4-1. Baseline data will be collected for a period of one year prior to IDU at each site where possible. In-use data will be collected for a period of one year following PCA. Between IDU and PCA, URET operations will be observed and trends in the metrics reported in order to understand any "learning curve" effects.

At ZID and ZME, URET prototype sites, URET will have been in daily use for quite some time, thus reducing the risk of URET CCLD. The prototype was built in increments. The prototype moved from a DYSIM environment to the operational floor at selected areas in January 1998. The number of URET-equipped sectors expanded to the entire center (both ZID and ZME) over the next year. Use of 2-way communications between Host and the URET prototype only began in July 1999. So, although much data prior to URET CCLD has been collected, it may not be possible to collect pure baseline data for these sites. Nevertheless, operations will continue to be observed at all URET CCLD sites, and sample data will be collected periodically in order to understand any trends relating to URET usage.

The IDU and PCA dates below reflect URET Build 1. URET Build 2 IDU (not shown below) is approximately 1 year after Build 1 IDU.

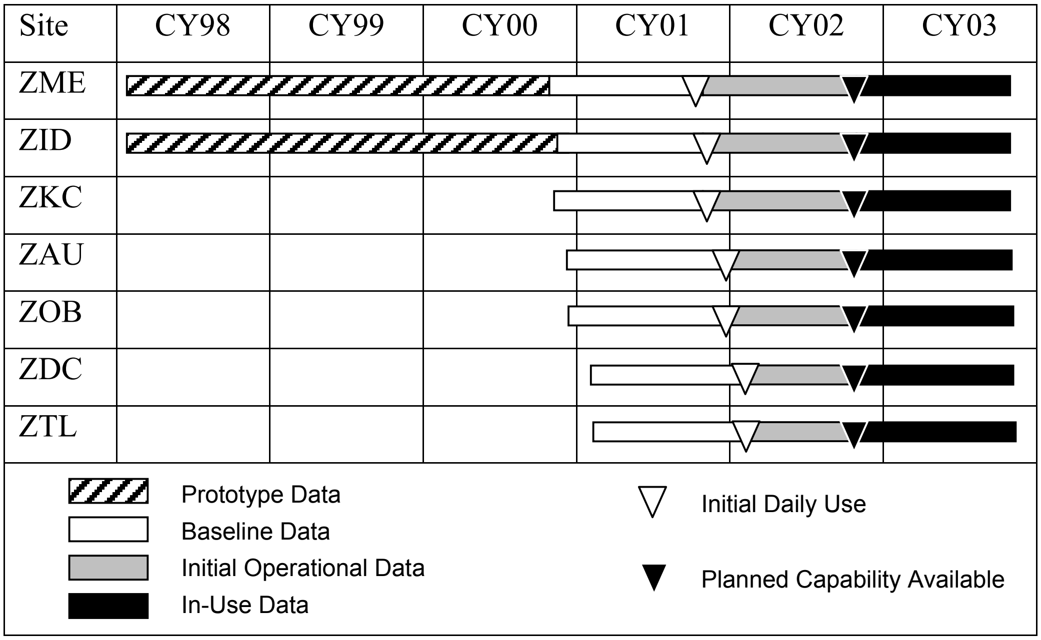


Figure 4-1. URET Evaluation Schedule

4.1.4 Data Collection

Assessing benefits requires the collection and manipulation of various types of data. Once the appropriate data are collected, they must be reduced into appropriate formats for analysis. Data requirements include information and run analyses on all flights, as well as sets of flights with common city-pair origin and destination, and do certain metrics and analyses on those subsets. Commonalties like same-day-of-week are also worth considering. Grouping by winds conditions will be investigated (as a proxy for the more general weather conditions). The following paragraphs identify possible data sources and related metrics as well as requirements for collecting the necessary data. These requirements pertain to operations in URET sectors in an ARTCC.

4.1.4.1 Data Sources

Monitor, Test, and Recording Files or Host Interface Device Files

These files contain all the information sent from the Host to URET prior to any URET processing. Monitor, Test, and Recording (MTR) files are URET recordings of information such as track and flight plan data obtained from the Host Computer System (HCS) General Purpose Output Interface Unit. Host Interface Device (HID) files record similar information, but via a new interface which will become operational in mid-1999. The HID files contain a superset of MTR files since the HID files are sent via the Host 320 patch which supports 2-way URET. Prior to May 1999, MTR files are available; afterwards, HID files are available.

Data Log Files

DLOG files are URET files that contain information pertaining to URET actions such as predicted conflicts and trial plans. These DLOG files contain time-stamped data records for URET runs. Many events can be recorded in DLOG files, including controller interactions, algorithmic processing, conflict notifications, number of trial plans built, and URET workstation configurations. A standard set of records is always recorded. Additional data records are available to turn on either at URET startup or during runtime.

Voice Recordings/Sector Observation

Observing sector activity or recreating sector activity through the review of voice recordings is a useful but very time consuming method for identifying certain specific actions. The voice recordings contain all verbal communications at a sector position while the microphone was keyed (e.g., conversations with pilots and with other controllers). Sector Observations require dedicated personnel to manually record certain activities. Sector Observations have the advantage of having the ATC context available via the plan view display (PVD) and conversations with the sector staff.

Facility Records

ARTCCs maintain information pertaining to restrictions and separation violations. Restrictions are recorded in LOAs and SOPs. This restriction information details flows into other facilities (ARTCCs or TRACONs) or into other sectors. It can include altitude restrictions (e.g., all jet traffic from sector *xx* bound for airport *yy* cross boundary at FL 230), route restrictions (e.g., "pref" routes), and flow restrictions (e.g., miles-in-trail).

System Analysis and Recording Data

SAR data are recordings generated by the Host that contain all non-voice information including radar reports within the ARTCC, and all flight plan messages processed by the Host. SAR tapes are converted into a useable format using the Data Analysis and Reduction Tool (DART). SAR data is needed when obtaining information from a location that does not have URET or the Host patch for the HID.

SAR contains very detailed information on Host processing for the entire center (e.g., the text of each datablock at each sector at every radar update (12 seconds)). SAR is recorded onto square tapes, and a single day at an ARTCC requires 30-50 SAR tapes for recording. For analysis, these tapes must be copied before their 2-week recycling deadline. These copied tapes must then be sent to another location for the DART extraction (e.g., the FAA Technical Center in New Jersey). Thus, analyzing large amounts of SAR data can quickly become a logistical concern.

Enhanced Traffic Management System Data

ETMS receives data from the host computer system. This information includes flight plan, flight amendment, cancellation, departure, ARTCC boundary crossing, sector assignment status, track data (position update), and arrival data. ETMS position reports are not as frequent as host track updates. Nevertheless, ETMS data may be useful in collecting information for metrics that do not require frequent position updates.

Rapid Update Cycle Winds

The Rapid Update Cycle (RUC) is an operational atmospheric prediction system comprising primarily of a numerical forecast model and an analysis system to initialize that model. The RUC has been developed to serve users needing short-range weather forecasts, including those in the US aviation community. RUC wind forecasts are used in URET trajectory modeling and can also be used to calculate an aircraft's true airspeed required for the air distance metric.

4.1.4.2 Data Sources for Each Metric

Table 4-2 contains a summary of the metrics with required data elements and sources.

Table 4-2. URET Metrics, Data Elements, and Data Sources

Outcome Category	Metric	Data Elements	Frequency	Data Source
Safety	Change in operational errors while capability is in use	Operational errors	per center	Facility records
	Change in operational deviations while capability is in use	Operational deviations	per center	Facility records
Delay/ Efficiency	Average enroute time and distance flown (on-time departures)	Aircraft positions and times	per flight	ETMS
	Average enroute air distance flown (on-time departures)	Wind vectors by position and time Aircraft positions and times	per operation	RUC, ETMS
	Average fuel usage	Fuel usage	per flight	Airlines
	Percentage of time spent at or near desired altitude for city pairs	Planned altitude	per flight	MTR/HID
		Actual altitudes flown	per flight	MTR/HID
	Number of restrictions eliminated - dynamic	Route restrictions Altitude restrictions	per center	SAR, MTR/HID, DLOG
	Number of restrictions eliminated - static	Route restrictions Altitude restrictions	per center	Adaptation files
	Aggregate degrees turned	Aircraft positions in center	per flight	MTR/HID
Predictability	Average planned versus actual enroute time and distance flown in center	Planned route	per flight	DLOG
		Actual route flown	per flight	DLOG
Flexibility	Average enroute time and distance flown (late departures)	Aircraft positions and times	per flight	ETMS
	Average enroute air distance flown (late departures)	Wind vectors by position and time Aircraft positions and times	per operation	RUC, ETMS
System Productivity	Number of aircraft per sector per unit time	Track control time	per sector	MTR/HID or DLOG
	M/A threshold	Monitor alert threshold values and alerts	per sector	Facility records

Issue: Many of the data sources provided represent FY99 activities. Sources for long term data collection are an open issue.

4.1.5 Evaluation Issues/Concerns

In addition to the issues and concerns listed with the individual metrics, there are issues that impact all the measures; most of these issues listed in this section deal with the dynamics of the ATC system.

Weather. It is accepted that weather has a major influence on the ATC system. In this discussion, weather encompasses both wind and precipitation. However, no accepted technique exists for quantifying weather, which is both qualitative and quantitative, and in any event is a multi-dimensional phenomenon. For example, there is no accepted technique to determine if all the Mondays in July had similar weather. Subjective opinions agree on extreme weather conditions (very good, very bad), but it's a difficult problem to analytically specify conditions that determine if two days under analysis have "similar" weather.

Change in ATC Resource. The NAS system is dynamic, and these measures are going to be collected over several years. The system will have changed over the years. For example, if a center makes a significant adjustment to a sector's boundaries during the data-collection period, the metrics will be comparing apples to oranges. It is possible that if such a change to the NAS is not considered during the evaluation impacts of the FFP1 capabilities will be exaggerated or not identifiable. Other NAS changes that will impact the metrics include adding runways or adjusting meter fixes.

Change in User Need of ATC Resources. Over the course of several years, the needs of the users could change, which in turn could significantly impact the ATC system. Examples of changes include the creation of new airlines or the formation of a new hub airport.

Aircraft Volume. Continued growth in the volume of air traffic is predicted during the data collection period. All metrics would need to be adjusted for the change in volume, but it is not always clear how to implement the adjustment. For example, if volume increased 5% and time in flight increased 6%, was a benefit provided by URET?

Agreement among Metrics. No metric accounts for all factors, and as such the metrics have to be considered all together as a collection of partial measures, all imperfect, of URET benefits. How to assess the potentially conflicting results of these metrics is unknown at this time.

4.2 Traffic Management Advisor – Single Center

4.2.1 Evaluation Overview

TMA is primarily used by traffic managers and air traffic controllers in the Air Route Traffic Control Centers (ARTCCs) to meter arrivals at large hub airports. Therefore our evaluation of TMA performance focuses on the contribution of the system to FAA performance outcomes within the expanded terminal airspace, and particularly within the ARTCC airspace immediately surrounding the particular TRACONs. The evaluation will examine changes in flying time, throughput, and fuel usage at the TMA airports attributable to TMA usage. Additionally, any changes in safety (as indicated by operational errors and deviations) will be analyzed and reported. Where possible data for each metric will be collected for a period of time prior to TMA implementation in order to establish baseline performance, and then for a period of time after TMA implementation so that the effect of TMA may be assessed. Local environmental, airport configuration, and airport demand data will also be collected in order to assess how these factors affect TMA usage and benefits, and in order to isolate the effects of TMA from those of changes in these conditions. Data will be collected for a period of one year prior to system deployment and one year following system deployment at each site so that seasonal factors may be fully removed, and so that any learning curve associated with TMA may be observed.

The TMA performance metrics are summarized in Table 4-3 on the following page, and each is described in the following sub-sections. For each metric we discuss the reason why the metric was selected, the direction in which we expect the metric to change following implementation of the particular tool, the data that will be required to calculate and understand the metric, and the data sources that we expect to rely on. At the end of this section we describe any concerns or issues related to computation or assessment of these metrics. All metrics will be computed for peak periods or during peak demand. A peak is defined as a period where the demand is close to or exceeds the AAR. The metrics are organized by the operational outcome categories to which they relate.

Table 4-3. TMA Performance Metrics

Outcome Category	Metric
Safety	Change in operational errors while capability is in use[a]
	Change in operational deviations while capability is in use[a]
User Access	Actual arrival rate
Delay/Efficiency	Mean flight time from 200 nmi range ring to meter fix[b]
	Mean arrival delay
	Mean fuel usage from 200 nmi range ring to meter fix[b]
	Variability of fuel usage from 200 nmi range ring to meter fix[b]
Predictability	Mean error in predicted meter fix arrival time
	Variability in error of predicted meter fix arrival time
	Variability of actual arrival rate
	Mean difference between airport acceptance rate and actual arrival rate
	Variability of time from 200 nmi range ring to meter fix[b]
System Productivity	Mean actual arrival rate/throughput per sector or position

[a] We intend to examine total operational errors/deviations for the airspace in which TMA operates. If there is any change in operational errors/deviations following TMA fielding we will attempt to determine if TMA has had any effect.
[b] A smaller range ring will need to be used for airports that are within 200 nmi of the ARTCC boundary.

4.2.2 Performance Metrics

4.2.2.1 Safety

The FFP1 core metrics that are intended to address system safety are the *change in operational errors* and the *change in operational deviations*. The FFP1 Program Office intends to track all operational errors and deviations that occur at each TMA site. An analysis will be performed to determine if there is a statistically significant increase in operational errors or deviations per arrival at each site after the implementation of TMA. Since TMA is not intended to directly improve or reduce system safety, it is expected that the number of operational errors/deviations per arrival will not change following TMA implementation, and it is desired that the number of operational errors/deviations per arrival will actually decrease. In addition to examining the overall number of operational errors/deviations and the number of operational errors/deviations per arrival, the cause of each operational error/deviation will be identified, and those that are either a direct or indirect consequence of TMA usage will be reported.

4.2.2.2 User Access

User access is described as the ability of users to act on and obtain services on demand. Measures of access try to quantify the quality and level of service, as well as the availability of system resources. These resources include physical resources (airports and airspace) and information resources.

Actual Arrival Rate

The metric that will be used to assess TMA's contribution to achieving this user-based outcome is the airport *actual arrival rate*. By helping controllers to optimize arrival flows, TMA should help to increase the overall throughput achievable. Increasing the achievable throughput will allow airlines operating into busy hub airports to schedule more flights at peak times, thereby increasing their access to the airport. This metric is intended to capture the ability of the TRACON to sustain a particular AAR. The data will be analyzed in specific blocks of time (15 mintues, 30 minutes) to capture this effect.

Local meteorological conditions and the airport configuration will also need to be collected in order to understand the variations of this metric better. In addition, we intend to examine departure rates, because at many airports arrival rates and departure rates are coupled either directly through shared runways or indirectly through shared taxiways and ramp areas.

4.2.2.3 Delay/Efficiency

The FFP1 core metrics that are intended to capture TMA's contribution to delay reduction are *mean flight time from the 200 nmi range ring to the meter fix* and *mean arrival delay*. TMA is designed to assist air traffic controllers and air traffic managers with balancing (or metering) the arrival streams at major airports. TMA should therefore help decrease the average time of flight for aircraft arriving at these airports.

As described earlier, the arrival airspace at Dallas/Ft. Worth International Airport (DFW), one of the TMA CCLD sites, may be segmented into TRACON and ARTCC regions (see Figure 4-2). Based on current and projected traffic, an ARTCC Traffic Management Coordinator (TMC) will create a plan to deliver aircraft to the TRACON at a rate that fully utilizes the capacity of the TRACON and the destination airport, subject to safety constraints. This plan consists of sequences and scheduled times of arrival at the meter fixes, which are way points that lie on the boundary between the ARTCC and TRACON airspaces. ARTCC controllers will issue clearances to aircraft so that they cross these fixes at the scheduled times determined by the TMC. TMA assists the TMC and the controllers by optimally scheduling the times of arrival of aircraft at the meter fixes.

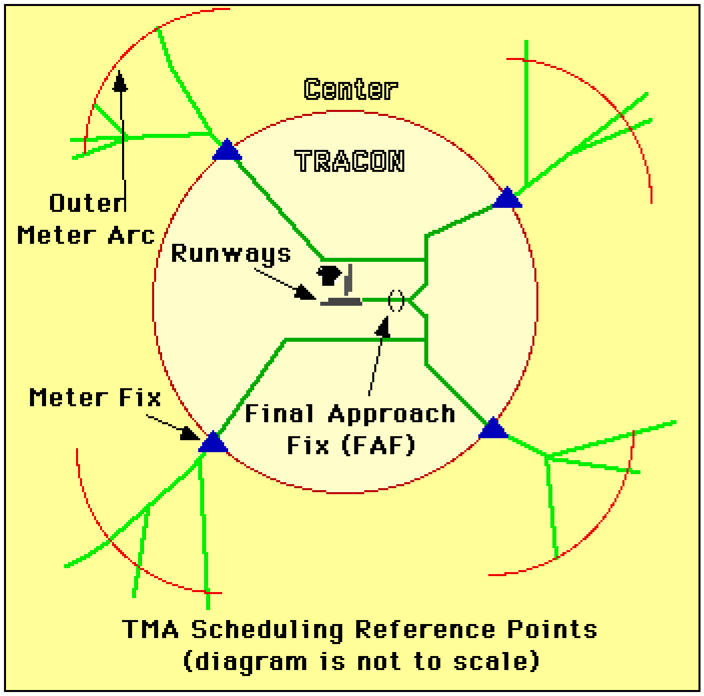


Figure 4-2. Notional Terminal Airspace

Mean Flight Time from 200nmi Range Ring to Meter Fix

TMA will typically begin to consider aircraft once they come within 200 nmi of the destination airport. At this point TMA will begin to compute an arrival sequence for the aircraft and a scheduled time of arrival at the meter fix. This arrival sequence and meter fix arrival time are recomputed continuously until the aircraft is estimated to be within 19 minutes of arriving at the specified meter fix. Once an aircraft reaches a meter fix it will be handed off to a TRACON controller, and TMA will cease to consider it. Therefore we feel that the most appropriate delay metric for TMA is the *mean flight time from the 200 nmi range ring to the meter fix*. This metric should capture the anticipated reduction in airborne holding in the extended terminal airspace. For those airports that are closer to the ARTCC boundary than 200 nmi a smaller range ring will need to be used, since single-center TMA cannot look into an adjacent ARTCC's airspace.

Mean Arrival Delay

Additionally, we will examine the *mean arrival delay* at the destination airport. Arrival delay is defined here as the difference between the scheduled arrival time (as published in the

Official Airline Guide) and the actual arrival time at the gate. Thus *mean arrival delay* is given by

$$\overline{D}_A = \frac{1}{n}\sum\left(t_{SchedArrival} - t_{ActArrival}\right)$$

Fuel Efficiency

It is possible that, through the use of TMA, arriving aircraft can be delayed by receiving vectors well before reaching the meter fixes while still at high altitudes. Prior to TMA introduction, these delays and holding would have occurred at lower altitudes. Thus while the flight time required to transit from the 200 nmi range ring to the meter fix may be unchanged from before TMA introduction, fuel usage may in fact decrease, since delays are being absorbed at higher, more fuel efficient altitudes. The fuel efficiency metrics are intended to capture this potential TMA benefit. The two fuel efficiency metrics are the *mean fuel usage from the 200 nmi range ring to the meter fix* and the *variability of fuel usage from the 200 nmi range ring to the meter fix*. Variability of fuel usage will be calculated as the standard deviation of fuel usage. For airports that are closer than 200 nmi to the ARTCC boundary smaller range rings will be used. Airport configuration, weather, and equipment will clearly be required data inputs in order to utilize these metrics. The airlines will play a vital role in supplying the necessary data to quantify this metric.

In order to better understand the performance of TMA under varying environmental conditions (for example, inclement weather), and also to separate the effects of these conditions on the metric from the effects of TMA itself, data relating to the local environment for each site will also be collected. For this metric it is anticipated that the airport configuration (e.g., north or south flow), ceiling, visibility, precipitation rate, and winds aloft might be collected.

4.2.2.4 Predictability

Predictability metrics are intended to measure the variation in the ATM system as experienced by the user. We have identified five predictability metrics for TMA:

- *Mean error in predicted meter fix arrival time*
- *Variability in error of predicted meter fix arrival time*
- *Variability of actual arrival rate*
- *Mean difference between airport acceptance rate and actual arrival rate*
- *Variability of flight time from 200 nmi range ring to meter fix*

Mean Error in Predicted Meter Fix Arrival Time

The first metric, *mean error in predicted meter fix arrival time*, is less a measure of user benefits than it is an indicator of the performance of the TMA system itself. As described earlier, TMA will provide a scheduled time of arrival at the meter fix for all arriving aircraft. This metric will provide an indication of the accuracy with which controllers are able to meter aircraft to the meter fixes as scheduled by TMA (there will not be any baseline value for this metric). The metric will be computed as follows:

$$\overline{E}_A = \frac{1}{n}\sum\left(t_{SchedMFArrival} - t_{ActMFArrival}\right)$$

TMA will continually update the scheduled arrival time of an aircraft at the meter fix until the "freeze horizon," when the aircraft is a specified distance or time from the meter fix (this distance or time is dependent on the type of aircraft, the particular meter fix, and the origin of the aircraft). For computing this metric the scheduled meter fix arrival time will be that value computed at the freeze horizon.

Variability in Error of Predicted Meter Fix Arrival Time

Another TMA predictability metric is the *variability in error of the predicted meter fix arrival time*. This metric will give an indication of the distribution of TMA errors and will be computed as follows:

$$V_{E_A} = \sqrt{\frac{\sum_{i=1}^{n}\left[\left(t_{SchedMFArrivali} - t_{ActMFArrivali}\right) - \overline{E}_A\right]^2}{n-1}}$$

Variability of Actual Arrival Rate

The next TMA predictability metric is the *variability of actual arrival rate*. This metric will provide an indication of how well TMA helps controllers to smooth out arrivals at the airport in question. We would expect the variability of the arrival rate to decrease after TMA implementation, while the mean arrival rate increases. Variability of arrival rate will be computed as follows:

$$V_{Arriv} = \sqrt{\frac{\sum_{i=1}^{n}\left[Arriv_i - \frac{1}{n}\left(\sum_{i=1}^{n} Arriv_i\right)\right]^2}{n-1}}$$

Mean Difference Between Airport Acceptance Rate and Actual Arrival Rate

We also will be reporting on the *mean difference between airport acceptance rate and actual arrival rate*, as we do for pFAST. TRACON managers will set a desired maximum aircraft arrival rate based on local conditions. This Airport Acceptance Rate (AAR) may be changed several times during the day, as meteorological and other conditions change. The difference between the AAR and the actual aircraft arrival rate is an indication of the excess capacity available. The analysis will focus on peak arrival periods, when high arrival demand levels necessitate arrival rates near the AAR. Variability will be estimated as the standard deviation of the difference between the AAR and the actual arrival rate during peak periods, i.e.,

$$\overline{\Delta}_{AAR/Arriv} = \frac{1}{n}\sum_{i=1}^{n}\left(Arriv_i - AAR_i\right)$$

Variability of Flight Time from 200nmi to Meter Fix

Local meteorological conditions will also be collected in order to better understand this metric. The *variability of flight time from the 200 nmi range ring to the meter fix* will also be computed to give an indication of the change in predictability of operations resulting from TMA implementation. Again, variability will be represented by the standard deviation, i.e.,

$$V_{Time} = \sqrt{\frac{\sum_{i=1}^{n}\left[t_i - \frac{1}{n}\left(\sum_{i=1}^{n} t_i\right)\right]^2}{n-1}}$$

4.2.2.5 Productivity

Finally, the FAA aims to increase airspace system productivity. Productivity is generally defined as the level of output achieved (in this case, air traffic services) per unit of input employed. Typically, productivity is measured with respect to input labor or capital, but other production inputs may be used. The FFP1 Program Office has elected to measure productivity in terms of NAS resources. Specifically, for the TMA system productivity will be measured in terms of the *mean actual arrival rate or throughput per sector or position*. This metric will be reported in two manners. First, the airport's actual arrival rate will be divided by the number of sectors adjacent to the TRACON feeding traffic to the airport. Second, the throughput of each of the meter fixes will be examined. This metric should provide an indication of the productivity of the sectors (and their corresponding controller positions) that are feeding traffic to the TMA destination airport. Again, local meteorological conditions and the airport configuration will be needed to understand the variations of this metric.

4.2.3 Evaluation Schedule

The TMA evaluation schedule is illustrated in Figure 4-3. Baseline data will be collected for a period of one year prior to Initial Daily Use (IDU) at each site where possible. In-use data will be collected for a period of one year following Planned Capability Available (PCA). Between IDU and PCA, TMA operations will be observed and trends in the metrics reported in order to understand any "learning curve" effects. At Dallas-Ft. Worth International a Build 1 version of TMA has been operational since the airspace surrounding the airport was redesigned in 1996. Therefore, it will not be possible to collect any meaningful baseline data for this site. Nevertheless, operations will continue to be observed at the Fort Worth ARTCC (ZFW) serving DFW airport, and sample data will be collected periodically in order to understand any trends relating to TMA usage.

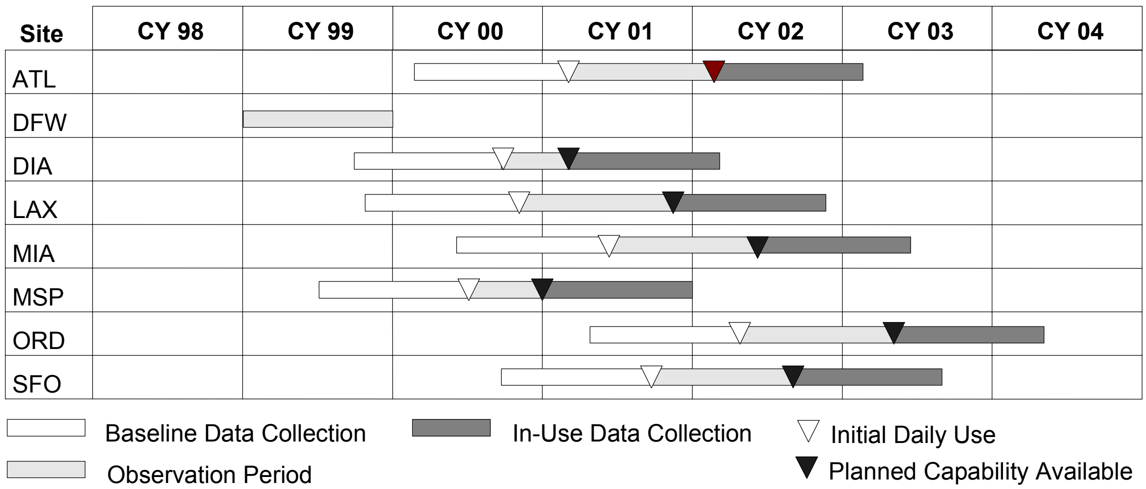


Figure 4-3. TMA Evaluation Schedule

4.2.4 Data Collection

A great deal of data will need to be collected if all of the metrics and associated factors described above are to be analyzed for a period of two years at each site. The TMA metrics and associated data elements are outlined in Table 4-4, along with the sources for these data elements (for those data elements to be collected by the FAA) and the frequency with which they will need to be collected. Each metric is discussed in more detail below (note that some of the metrics have been grouped together as they rely on the same data sources and will be evaluated using the same context data).

Operational errors/operational deviations - these metrics will be derived from the National Aviation Safety Data Analysis Center (NASDAC) database and facility records.

Mean/variability of flight time from 200 nmi range ring to meter fix - these metrics will be derived from aircraft track messages obtained from ETMS or ARTS. For ZFW, NASA has collected data from the prototype TMA system there, so TMA recorded data could be used. Associated data may need to be collected from ETMS or ARTS, and hourly surface weather data are collected and distributed by the National Climatic Data Center.

Mean arrival delay - this metric will be derived from ASQP (actual arrival time) and OAG (scheduled arrival time) data bases. Actual arrival rate will be collected from either

ETMS, ARTS or CTAS recorded data. Hourly surface weather observations are available from the NCDC.

Mean/variability of fuel usage from 200 nmi range ring to meter fix - Fuel usage and associated flight information (i.e., flight number, origin city, equipment) will need to be provided by the airlines. Arrival and departure rate information may be obtained from ETMS, ARTS or from CTAS recorded data. Hourly surface weather observations are available from the NCDC.

Mean/variability of error in predicted meter fix arrival time - the scheduled meter fix arrival time at the freeze horizon and the actual meter fix arrival time will be obtained from TMA recorded data.

Mean difference between airport acceptance rate and actual arrival rate, mean/variability of actual arrival rate - The AAR is recorded on TRACON logs, and this data is currently being collected for DFW. Actual arrival rates are also recorded on TRACON logs, or they may be obtained from ETMS or ARTS. Hourly surface weather observations are available from the NCDC.

Mean actual arrival rate, mean actual arrival rate/throughput per sector or position - Actual arrival rate will be obtained from ETMS, ARTS or TRACON logs. Meter fix throughput may be obtained from ETMS or ARTS data. The number of relevant sectors and positions may be obtained from a site survey. Hourly surface weather observations are available from the NCDC.

Table 4-4. TMA Metrics, Data Elements, and Data Sources

Outcome Category	Metric	Data Elements, *Local Factors*	Frequency	Data Source
Safety	Change in operational errors while capability is in use	Operational error report	on occurrence	NASDAC, Facility records
	Change in operational deviations while capability is in use	Operational deviation report	on occurrence	NASDAC, Facility records
User Access	Mean actual arrival rate	Arrival rate	30 minutes	ETMS or TRACON logs
		Ceiling	hourly	NCDC
		Visibility	hourly	NCDC
		Precipitation rate	hourly	NCDC
Delay/ Efficiency, Predictability	Mean flight time from 200nmi range ring to meter fix, Variability of flight time from 200nmi range ring to meter fix	Time from 200nmi range ring to MF	per operation	calculated
		Aircraft x,y position	4 minutes	ETMS
		Arrival fix	per operation	TMA or derived
		Call sign	per operation	ETMS
		Actual arrival time	per operation	ETMS
		Equipment type	per operation	ETMS
		Arrival rate	30 minutes	TMA or ETMS
		Departure rate	30 minutes	ETMS
		Ceiling, visibility, precipitation rate	hourly	NCDC
Delay/ Efficiency	Mean arrival delay	Arrival delay	per operation	calculated
		Airport	per operation	ASQP
		Call sign	per operation	ASQP
		Actual arrival time	per operation	ASQP
		OAG arrival time	per operation	OAG
		Arrival rate	30 minutes	TMA or ETMS
		Departure rate	30 minutes	ETMS
		Ceiling, visibility, precipitation rate	hourly	NCDC
Delay/ Efficiency	Mean fuel usage from 200nmi range ring to meter fix	Fuel usage	per operation	Airlines
		Call sign	per operation	Airlines
		Departure city	per operation	Airlines
		Actual arrival time	per operation	Airlines
		Equipment type	per operation	Airlines
		Arrival rate	30 minutes	TMA or ETMS
		Departure rate	30 minutes	ETMS
		Ceiling, visibility, precipitation rate	Hourly	NCDC
Delay/ Efficiency, Predictability	Mean error in predicted meter fix arrival time, Variability in error of predicted meter fix arrival time	Error in meter fix arrival time	per operation	Calculated
		Initial estimated mf arrival time	per operation	TMA
		Actual mf arrival time	per operation	TMA
		Call sign	per operation	TMA
		Actual arrival time	per operation	TMA
		Equipment type	per operation	TMA
		Arrival rate	30 minutes	TMA
		Arrival fix	per operation	TMA
Delay/ Efficiency, Predictability	Mean difference between AAR and actual arrival rate, Variability of actual arrival rate	Acceptance rate	when changed	TRACON logs
		Arrival rate	30 minutes	ETMS or TRACON logs
		Ceiling, visibility, precipitation rate	hourly	NCDC
		Equipment type	per operation	TMA
		Arrival rate	30 minutes	TMA
		Arrival fix	per operation	TMA
System Productivity	Mean actual arrival rate/throughput per sector or position	Arrival rate	30 minutes	ETMS or TRACON logs
		Number of arrival sectors	Once	Site survey
		Number of arrival positions	Once	Site survey
		Ceiling, visibility, precipitation rate	hourly	NCDC

4.2.5 Evaluation Issues/Concerns

Safety

The metrics being used to capture changes in safety resulting from TMA usage are the numbers of operational errors and operational deviations. Since there are only a very small number of operational errors or deviations at any given FAA facility during any month, it will likely be difficult to discern any statistically significant differences in this metric.[2] Furthermore, it may prove difficult to determine if a particular operational error was either resulting from or exacerbated by TMA usage.

User Access

The TMA access metric is the *mean actual arrival rate*. While this metric seems to capture the users' ability to access the NAS, we may not see any changes in the metric until airlines adjust to the new and presumably improved operating environment by changing their schedules accordingly.

Delay/Efficiency

One of the TMA delay metrics is the *mean flight time from the 200 nmi range ring to the meter fix*. All things being equal, we would expect this mean flight time to decrease following TMA fielding. However, care must be taken to account for varying arrival routings, airport configurations, and arrival demand. Weather effects can especially impact flight times. In particular, convective activity in the vicinity of an airport can significantly increase delays, and there is currently no database available that captures such severe weather for all of the FFP1 sites.

In order to understand this metric better, we intend to segment the flight time into two components. At many ARTCCs, controllers will first meter TRACON-bound aircraft to an outer meter arc, which is a fixed distance from the meter fix (see Figure 4-2). From this meter arc they will then meter aircraft to the meter fix. In addition to examining the mean flight time from the 200 nmi range ring to the meter fix, we will also analyze the time from the 200 nmi range ring to the outer meter arc, and the time from the outer meter arc to the meter fix. In this way we hope to better understand exactly how TMA is assisting controllers in metering airport arrivals.

Our second TMA delay metric is *mean arrival delay*. As mentioned earlier, we intend to compute this metric by comparing actual gate arrival times with published scheduled arrival

[2] For example, there are typically about 0.99 operational errors per 100,000 operations in ARTCC airspace. For an airport with 1 million operations annually, one would expect 9.9 errors in ARTCC airspace.

times from the OAG. The difficulty in examining trends in arrival delay is that scheduled times generally tend to increase over time, as airline schedulers respond to increasing delay in the system. Furthermore, it is well known that airlines typically pad block times in order to improve on-time performance. It may therefore prove difficult to draw any conclusions regarding TMA's effects on arrival delay, given the continual adjustments being made to schedules as well as the continuous evolution of the NAS. Nevertheless, airlines place high value on meeting their published schedules. Because of the hub-and-spoke system favored by most airlines, a late arrival at a hub can result in major disruptions of the airline's networ for the remainder of the day. For this reason we intend to examine delay relative to the schedule as well as the differences in actual flight times. In order to isolate any arrival delay changes caused by TMA from upstream (e.g., departure or enroute) delays that are not affected by TMA, we will also examine these delays if warranted.

The fuel efficiency metrics are *mean and variability of fuel usage from the 200 nmi range ring to the meter fix*. It will be necessary that the airlines provide this detailed data to accurately quantify these metrics.

Predictability

Two related predictability metrics for TMA are the *variability of arrival rate* and the *mean difference between airport acceptance rate and actual arrival rate*. These metrics may be more an indication of user access to the NAS than predictability of the NAS, as we mentioned earlier. Furthermore, it may be difficult to mathematically define a peak period.

Productivity

The TMA productivity metric is *mean arrival rate/throughput per sector or position*. This metric is extremely similar to that for access, the *mean actual arrival rate*. We do not expect there to be any changes in the design of ARTCC airspace or the size of the ARTCC staff resulting from the implementation of TMA. Therefore, we may not observe any changes to this metric until airlines adjust their schedules to the new operating environment.

4.3 Passive Final Approach Spacing Tool

4.3.1 Evaluation Overview

As pFAST is primarily used by traffic managers and air traffic controllers in TRACONs, our evaluation of pFAST performance focuses on the contribution of pFAST to FAA performance outcomes within the TRACON airspace. The evaluation will examine changes in flying time and throughput at the pFAST airports attributable to pFAST usage. Additionally, any changes in safety (as indicated by operational errors and operational deviations) will be analyzed and reported. Where possible data for each metric will be collected for a period of time prior to pFAST implementation in order to establish baseline performance, and then for a period of time after pFAST implementation so that the effect of pFAST may be assessed. Local environmental, airport configuration, and airport demand data will also be collected in order to assess how these factors affect pFAST usage and benefits, and in order to isolate the effects of pFAST from those of changes in these conditions. Data will be collected for a period of one year prior to system deployment and one year following system deployment at each site so that seasonal factors may be fully removed, and so that any learning curve associated with pFAST may be observed.

The pFAST performance metrics are summarized in Table 4-5 on the following page, and each is described in the following sub-sections. For each metric we discuss the reason why the metric was selected, the direction in which we expect the metric to change following implementation of the particular tool, the data that will be required to calculate and understand the metric, and the data sources that we expect to rely on. At the end of this section we describe any concerns or issues related to computation or assessment of these metrics. All metrics will be computed for peak periods or peak demand, where a peak is defined as a period where the demand is close to or exceeds the AAR. Departures will also be examined in order to determine if pFAST has had any effect on departure rates. The metrics are organized by the operational outcomes to which they relate.

Table 4-5. pFAST Performance Metrics

Outcome Category	Metric
Safety	Change in operational errors while the capability is in use[a]
	Change in operational deviations while the capability is in use[a]
User Access	Actual arrival rate
	Actual arrival rate for each runway
Delay/Efficiency	Mean flight time from meter fix to runway threshold
	Mean fuel usage from meter fix to threshold
	Variability of fuel usage from meter fix to threshold
Predictability	Mean difference between airport acceptance rate and actual arrival rate
	Variability of flight time from meter fix to runway threshold
System Productivity	Distribution and throughput of operations per runway/position

[a] We intend to examine total operational errors/deviations for the airspace in which pFAST operates. If there is any change in operational errors/deviations following pFAST fielding we will attempt to determine if pFAST has had any effect.

4.3.2 Performance Metrics

4.3.2.1 Safety

The FFP1 core metrics that are intended to address system safety are the *change in operational errors* and the *change in operational deviations*. The FFP1 Program Office intends to track all operational errors and operational deviations that occur at each pFAST site. An analysis will be performed to determine if there is a statistically significant increase in operational errors or deviations per arrival at each site after the implementation of pFAST. Since pFAST is not intended to directly improve or reduce system safety, it is expected that the number of operational errors and deviations per arrival will not change following pFAST implementation, and it is desired that the number of operational errors and deviations per arrival will actually decrease. In addition to examining the overall number of operational errors/deviations and the number of operational errors/deviations per arrival, the cause of each operational error/deviation will be identified, and those that are either a direct or indirect consequence of pFAST usage will be reported.

4.3.2.2 User Access

User access is described as the ability of users to act on and obtain services on demand. Measures of access try to quantify the quality and level of service, as well as the availability of system resources. These resources include physical resources (airports and airspace) and information resources. The metrics that will be used to assess pFAST's contribution to achieving this outcome are the *actual arrival rate* and the *arrival rate for each runway*. By evening out the flows between arrival runways, pFAST should help to increase the overall throughput achievable. Increasing the achievable throughput will allow airlines operating into busy hub airports to schedule more flights at peak times, thereby increasing their access to the airport. Although in actual operations, access may be more limited by the number of gates than by the peak number of aircraft passed through, this metric deliberately singles out runway throughput over access to gates. Furthermore, although we might see the mean arrival rate for a particular runway decrease as a result of pFAST usage, we would expect to see the overall mean for a given airport increase. This might happen if the runways had not previously been well balanced and this particular runway was being over-assigned.

This metric is also intended to capture the ability of the TRACON to sustain a particular AAR. The data will be analyzed in specific blocks of time (15 mintues, 30 minutes) to capture this effect. Local meteorological conditions and the airport configuration will also need to be collected in order to understand the variations of this metric better.

4.3.2.3 Delay/Efficiency

Measures of system delay are intended to quantify the time savings (the amount of time beyond expectation) that it takes to complete an operation.

Mean Flight Time from Meter Fix to Threshold

The FFP1 performance metric that is intended to capture pFAST's contribution to system delay reduction is *mean flight time from the meter fix to the runway threshold*. pFAST is designed to assist air traffic controllers with balancing the arrival streams at airports with more than one runway. In this way pFAST should help decrease the average time of flight for aircraft arriving at the airport in question.

The arrival airspace at DFW, one of the pFAST CCLD sites, can be segmented into TRACON and enroute Center regions. Figure 4-4 illustrates the TRACON air space surrounding DFW, which extends approximately 40 nautical miles from the airport. Arrival traffic is merged at four "meter fixes" on the Center-TRACON boundary. These meter fixes correspond to the four primary arrival directions (i.e., northeast, southeast, southwest, and northwest). During heavy traffic periods aircraft are funneled through these fixes, or "feeder gates," as a means of controlling or metering the flow rate into the terminal area. Jet and propeller aircraft flowing into each meter fix are separated into two independent streams vertically separated by 2,000 feet, thus helping controllers to avoid overtaking conflicts between aircraft with significantly different approach speeds. pFAST "looks" at all aircraft within the TRACON airspace, and determines the near-optimal sequencing of these aircraft to the active arrival runways in order to minimize delay. Thus the mean flight time from the meter fixes to the runway threshold for all arriving aircraft seems to be an appropriate metric with which to gauge pFAST performance.

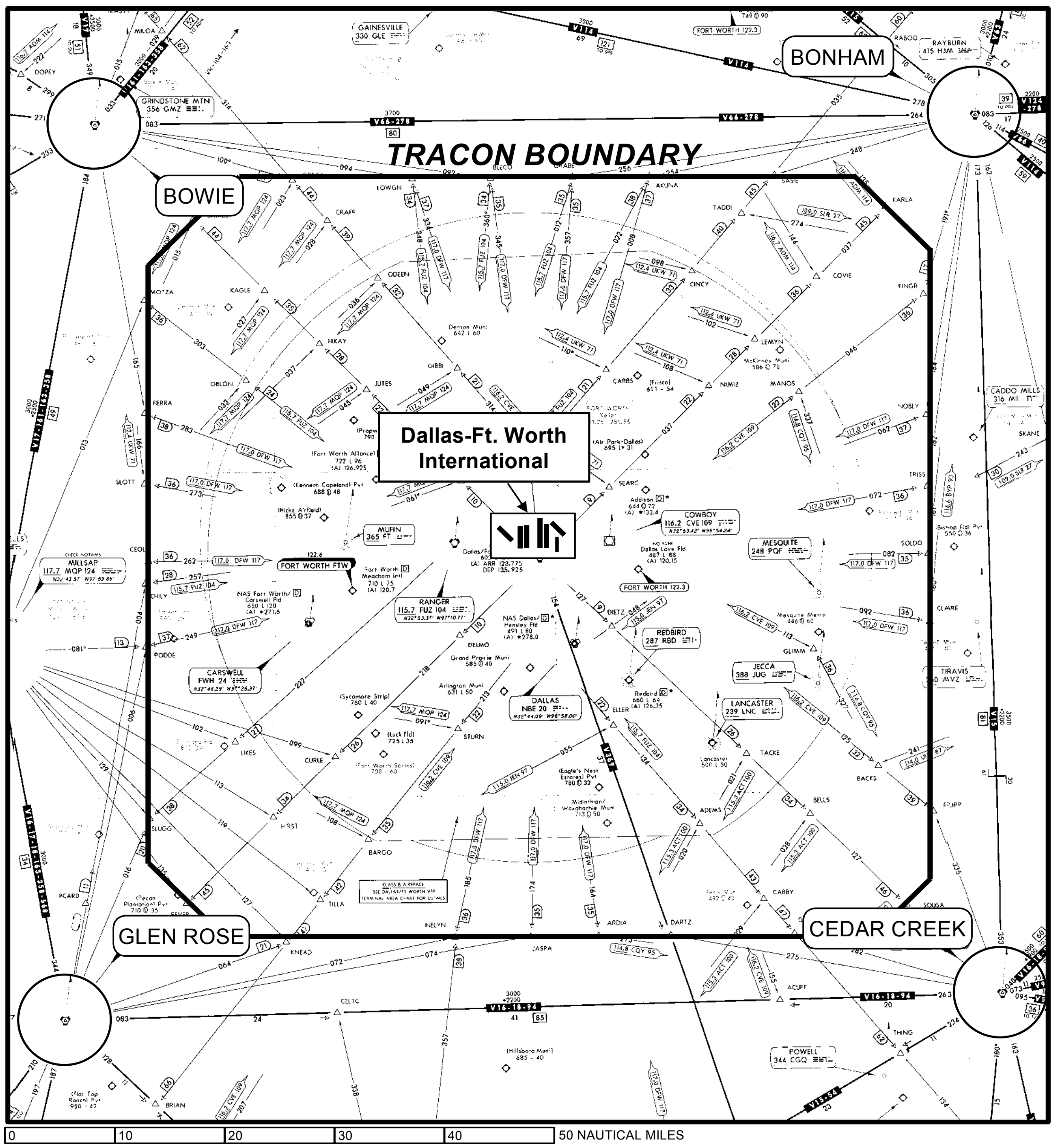


Figure 4-4. DFW TRACON Airspace and Meter Fixes

In addition to evaluating this metric subsequent to fielding of pFAST at each of the sites, data will be collected prior to the start of operational use of the tool. This baseline data will allow a comparison to be made of average flying times before and after pFAST implementation, and in this way the expected improvement in flying time at each site resulting from pFAST will be obtainable.

Mean Fuel Usage from Meter Fix to Threshold

The first pFAST fuel efficiency metric is the *mean fuel usage from the meter fix to the runway threshold*. As mentioned earlier, by smoothing out arrival flows pFAST should reduce aircraft delays in the terminal airspace. This reduction in delays will lead to a corresponding reduction in fuel usage. We also intend to examine the *variability of fuel usage from the meter fix to the runway threshold*. The variability of fuel usage is an indication of the ease with which airlines may optimize fuel loads and thereby minimize their fuel costs.

In order to better understand the performance of pFAST under varying environmental conditions (for example, inclement weather), and also to separate the effects of these conditions on the metric from the effects of pFAST itself, data relating to the local environment for each site will also be collected. For this metric it is anticipated that the airport configuration (e.g., north or south flow), ceiling, visibility, precipitation rate, and winds aloft might be collected. This data might then be used to aide in the evaluation of the metric in two distinct manners. In the first and simplest to understand technique, the data would be segmented into "similar" time periods. For example, dates would be selected that had similar weather conditions and airport demand, both before and after implementation of the tool. The metric would be calculated for these days, and the process would be repeated in order to generate a "picture" of how the tool affects the metric under a range of environmental and demand conditions. In a somewhat more mathematically sophisticated approach, a multiple regression model could be estimated for which the metric is the dependent variable and the environmental variables are the independent variables. In this approach all of the data may be used at once to give an indication of the effects of the tool, the environmental factors, and the interaction of the two on the metric.

4.3.2.4 Predictability

Predictability metrics are intended to measure the variation in the ATM system as experienced by the user. There are two FFP1 performance metrics that should capture the effects of pFAST on NAS predictability.

Mean Difference Between AAR and Actual Arrival Rate

The first predictability metric is the *mean difference between the airport acceptance rate and actual arrival rate.* Traffic Management supervisors will set a desired maximum AAR based on local conditions. This AAR may be changed several times during the day, as meteorological and other conditions change. The difference between the AAR and the actual aircraft arrival rate is an indication of the excess capacity available. Figure 4-5 illustrates the variability of the arrival rate and AAR during a one week period at DFW (arrival rates are based on 30 minute periods). The analysis will focus on peak arrival periods, when high arrival demand levels necessitate arrival rates near the AAR. Variability will be estimated as the standard deviation of the difference between the AAR and the actual arrival rate during peak periods, i.e.,

$$\overline{\Delta}_{AAR/Arriv} = \frac{1}{n}\sum_{i=1}^{n}\left(Arriv_i - AAR_i\right)$$

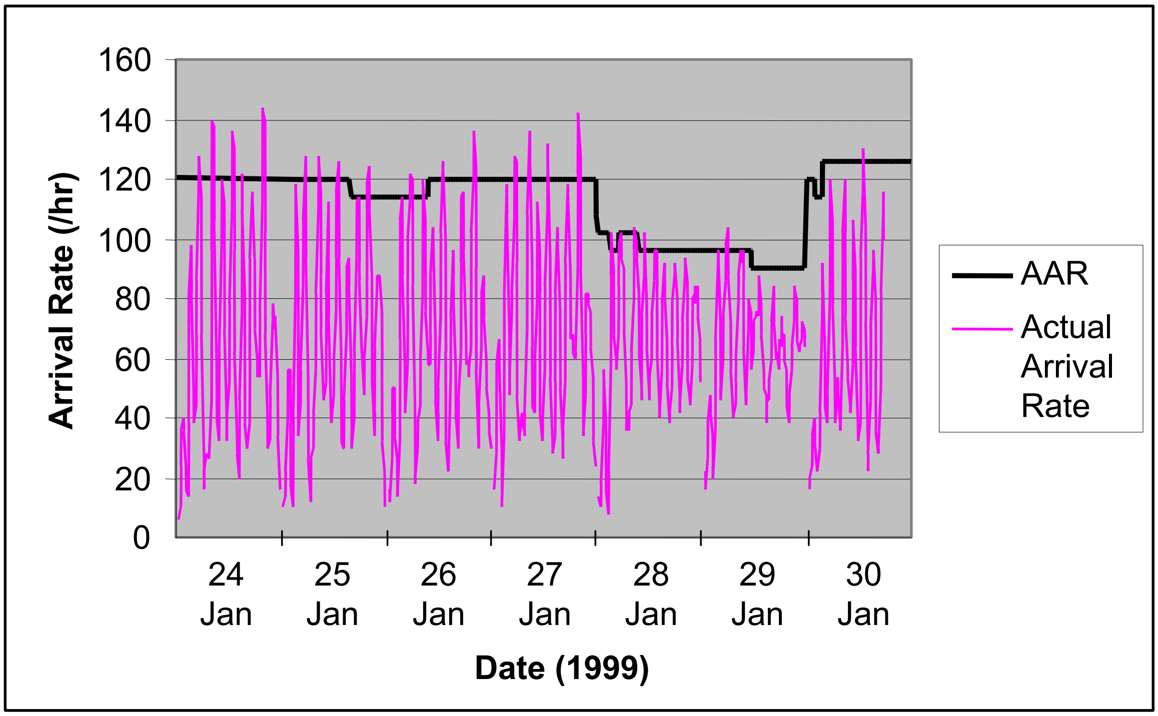


Figure 4-5. Representative Arrival Rates, DFW

Variability of Flight Time from Meter Fix to Threshold

The second pFAST predictability metric is the *variability of flight time from the meter fix to the threshold.* This metric provides an indication of the consistency in flight times that airlines can expect in the terminal area. Again, variability will be represented by the standard deviation, i.e.,

$$V_{Time} = \sqrt{\frac{\sum_{i=1}^{n}\left[t_i - \frac{1}{n}\left(\sum_{i=1}^{n} t_i\right)\right]^2}{n-1}}$$

4.3.2.5 Productivity

Finally, the FAA aims to increase airspace system productivity. Productivity is generally defined as the level of output achieved (in this case, air traffic services) per unit of input employed. Typically, productivity is measured with respect to input labor or capital, but other production inputs may be used. The FFP1 Program Office has elected to measure productivity in terms of NAS resources. Specifically, for the pFAST system the productivity of arrival runways will be measured in terms of the *distribution and throughput of operations per runway or position*. As mentioned earlier, pFAST is expected to improve the balance of arrivals on alternate runways, and a previous metric was the average arrival rate per runway. This metric examines the number of operations (to include departures) per *position*, and the time distribution of operations on each of the runways at a pFAST site. This distribution is expected to provide insights into how effective pFAST is in aiding controllers to balance the runways, or maximize the productivity of the runways. Again, local meteorological conditions and the airport configuration will be needed to understand the variations of this metric.

4.3.3 Evaluation Schedule

The pFAST evaluation schedule is illustrated in Figure 4-6. As mentioned earlier, baseline data will be collected for a period of one year prior to IDU at each site where possible. In-use data will be collected for a period of one year following PCA. Between IDU and PCA, pFAST operations will be observed and trends in the metrics reported in order to understand any “learning curve” effects. At DFW a Build 1 version of pFAST became operational in February 1999. Baseline data will be collected for a period of one year prior to this, and in-use data will be collected beginning in August 1999. Trends will be observed between February and August 1999.

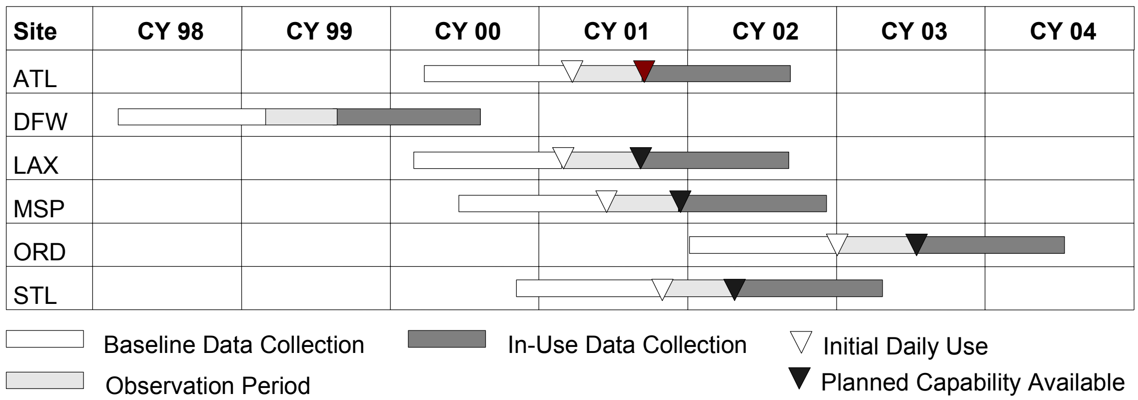


Figure 4-6. pFAST Evaluation Schedule

4.3.4 Data Collection

A great deal of data will need to be collected if all of the metrics and associated factors described above are to be analyzed for a period of two years at each site. The pFAST metrics and associated data elements are outlined in Table 4-6, along with the sources for these data elements (for those data elements to be collected by the FAA) and the frequency with which they will need to be collected. Each metric is discussed in more detail below (note that some of the metrics have been grouped together as they rely on the same data sources and will be evaluated using the same context data).

Table 4-6. pFAST Metrics, Data Elements, and Data Sources

Outcome Category	Metric	Data Elements, *Local Factors*	Frequency	Data Source
Safety	Change in operational errors while capability is in use	Operational error report	on occurrence	NASDAC, Facility records
	Change in operational deviations while capability is in use	Operational deviation report	on occurrence	NASDAC, Facility records
User Access, System Productivity	Mean arrival rate, Mean arrival rate per runway, Distribution and throughput of operations per runway/position	Arrival rate	30 minutes	ARTS or TRACON logs
		Ceiling	hourly	NCDC
		Visibility	hourly	NCDC
		Precipitation rate	hourly	NCDC
Delay/ Efficiency, Predictability	Mean flight time from meter fix to runway threshold, Variability of flight time from meter fix to runway threshold	Time from meter fix to threshold	per operation	calculated
		Aircraft x,y position	4 minutes	ARTS
		Arrival fix	per operation	pFAST or derived
		Call sign	per operation	ARTS
		Actual arrival time	per operation	ARTS
		Equipment type	per operation	ARTS
		Arrival rate	30 minutes	pFAST or ARTS
		Departure rate	30 minutes	ARTS
		Ceiling, visibility, precipitation rate	hourly	NCDC
	Mean difference between AAR and actual arrival rate	Acceptance rate	when changed	TRACON logs
		Arrival rate	30 minutes	ARTS or TRACON logs
		Ceiling, visibility, precipitation rate	hourly	NCDC
		Equipment type	per operation	ARTS
		Arrival rate	30 minutes	ARTS
		Arrival fix	per operation	ARTS
	Mean gate-to-gate fuel usage, Variability of gate-to-gate fuel usage	Fuel usage	per operation	Airlines
		Call sign	per operation	Airlines
		Departure city	per operation	Airlines
		Actual arrival time	per operation	Airlines
		Equipment type	per operation	Airlines
		Arrival rate	30 minutes	pFAST or ETMS
		Departure rate	30 minutes	ARTS
		Ceiling, visibility, precipitation rate	hourly	NCDC

Operational errors and deviations - these metrics will be derived from the National Aviation Safety Data Analysis Center (NASDAC) database and facility records.

Mean flight time/variability of flight time from Meter Fix to runway threshold - this metric will be derived from aircraft radar track messages obtained from the ARTS radar processor computers. For ZFW, NASA has collected these data from the prototype pFAST system there, so pFAST recorded data will be used. Associated data may need to be collected from ARTS, and hourly surface weather data are collected and distributed by the National Climatic Data Center.

Mean difference between airport acceptance rate and actual arrival rate - The AAR is recorded on TRACON logs. Hourly surface weather observations are available from the NCDC.

Mean/variability of fuel usage from meter fix to runway threshold - Fuel usage and associated flight information (i.e., flight number, origin, equipment) will need to be provided by the airlines. Arrival and departure rate information may be obtained from ARTS, CTAS recorded data, or TRACON logs. Hourly surface weather observations are available from the NCDC.

Distribution and throughput of operations per runway/position - Arrival rate for each runway is available in TRACON logs in ten minute intervals. Hourly surface weather observations are available from the NCDC.

4.3.5 Evaluation Issues/Concerns

Safety

The metrics being used to capture changes in safety resulting from pFAST usage are the numbers of operational errors and operational deviations. Since there are only a very small number of operational errors at any given FAA facility during any month, it will likely be difficult to discern any statistically significant differences in these metrics.[3] Furthermore, it may prove difficult to determine if a particular operational error was either resulting from or exacerbated by pFAST usage.

User Access

The pFAST access metrics are the *mean actual arrival rate* and the *mean actual arrival rate for each runway*. While these metrics seems to capture users' ability to access the NAS, we may not see any changes in the metrics until airlines adjust to the new and presumably improved operating environment by changing their schedules accordingly. When examining runway balancing we will need to consider any differences in lengths and/or capabilities of the arrival runways at a given airport.

Delay/Efficiency

The pFAST delay metric is the *mean flight time from the meter fix to the runway threshold*. All things being equal, we would expect this mean flight time to decrease following pFAST fielding. However, care must be taken to account for varying arrival routings, airport configurations, and arrival demand. Weather effects can especially impact flight times. In particular, convective activity in the vicinity of an airport can significantly increase delays, and there is currently no database available that captures such severe weather for all of the FFP1 sites.

---

[3] For example, there are typically about 0.81 operational errors per 100,000 operations in TRACON airspace. For an airport with 1 million operations annually, one would expect 8.1 errors in TRACON airspace.

While pFAST recorded data may be used to compute this metric once the system is installed at a site, ARTS data will be required to calculate the metric prior to system fielding (we believe that ETMS data will not prove accurate enough to discern small differences in this metric resulting from pFAST usage). The data are not currently available at all of the FFP1 sites.

The fuel efficiency metrics are the *mean and variability of fuel usage from the meter fix to the runway threshold*. This detailed data may be difficult to obtain.

Predictability

The first predictability metric for pFAST is the *mean difference between airport acceptance rate and actual arrival rate.* This metric, which can be computed using data available in TRACON logs, may be more an indication of user access to the NAS than predictability of the NAS. And while it is not difficult to calculate this metric, it may be difficult to mathematically define a peak period or peak demand.

System Productivity

The pFAST productivity metric is *distribution and throughput of operations per runway or position*. This metric is extremely similar to that for access, the *mean actual arrival rate*. We do not expect there to be any changes in the design of TRACON airspace, the number of positions, or the size of the TRACON staff resulting from the implementation of pFAST. Therefore, we may not observe any changes to this metric until airlines adjust their schedules to the new operating environment.

4.4 CDM: Enhanced Ground Delay Program

Collaborative Decision Making (CDM) was conceived out of the FAA's Airline Data Exchange (FADE) experiments that began in 1993. These experiments proved that having airlines send updated schedule information to the FAA could improve air traffic management decision making. CDM has evolved from these same principles in an effort to improve air traffic management through information exchange and data sharing.

The initial focus of CDM, known as Enhanced Ground Delay Program (GDP-E), started prototype operations at San Francisco (SFO) and Newark (EWR) airports in January 1998. Under GDP-E, participating airlines send operational schedules and changes to schedules to the Air Traffic Control Systems Command Center (ATCSCC) on a continuous basis. This schedule information includes, but is not limited to, flight delay information, cancellations, and newly created flights. The ATCSCC uses this information to better implement and manage ground delay programs (GDPs).

4.4.1 Evaluation Overview

In August 1998, The National Center of Excellence for Aviation Operations Research (NEXTOR) released a benefits assessment on CDM. This preliminary report highlights several areas that have produced tangible benefit as well as areas that are prospects for future benefits. It later proved to be a valuable resource to the development of a thorough GDP-E evaluation plan. As a result, several of the metrics noted in this evaluation plan have been either derived or directed from the NEXTOR report.

In February 1999, the CDM Working Group (including AOZ, ASD-400, Metron, Inc., MITRE/CAASD, NEXTOR, SETA, and Volpe) began identifying metrics and developing methodologies to identify operational impacts of GDP-E and its components on the NAS. Since the fundamental principal underlying GDP-E (and CDM) is the improvement of information and the sharing of data, it is important for the GDP-E section of the evaluation plan to incorporate the knowledge, ideas, and experience of all groups that are functionally involved.

It should be noted that this Metrics Plan and the subsequent evaluation will build from the efforts and direction provided by the CDM Working Group and the NEXTOR report. It is not intended to supplement *their* ongoing efforts.

Based on evidence that TFM (Traffic Flow Management) constraint information is being distributed more efficiently and reliably, NAS users can expect a reduction in system delays. Therefore, our evaluation of GDP-E performance will examine changes in flight time, taxi-out time, distance flown, and the number of airport operations. Since GDP-E is beyond prototype operations and therefore further along in evaluation process, operational assessments are underway. Results obtained from these assessments, using the metrics identified in this Metrics Plan, will be documented in a preliminary report due to be released at the end of September 1999. Several of the metrics being analyzed and those originally identified by the RTCA Metrics Working Group, have been eliminated because of problems associated with a pre-GDP-E and post-GDP-E analysis. Such problems include, but are not limited to, normalization, interpretability, and data availability.

Although GDP-E is currently being used by Airline Operation Centers (AOCs) *and* the Air Traffic Control System Command Center (ATCSCC) the evaluation will focus primarily on GDP-E performance to the non-FAA users. This is not to say that non-FAA users will be the only beneficiaries of GDP-E impacts, but in fact all users of the NAS will likely realize improved performance.

The GDP-E performance metrics are summarized in Table 4-7 on the following page.

Table 4-7. GDP-E Performance Metrics

Outcome Category	Metric
Safety	Change in operational errors while capability is in use[a] Change in operational deviations while capability is in use[a]
User Access	Number of unused slots
Delay/Efficiency	Mean flight time (excess arrival flying time)
	Compression minutes saved
Predictability	Integrated Predictive Error (IPE)
	Rate Control Index (RCI)
	EDCT compliance ratio
	Number of GDPs canceled near start
	Number of GDP revisions
Flexibility	Mean distance flown
	Control Time of Arrival
Additional Metrics [b]	Mean taxi-out time
	Variability in distance flown
	Variability in planned arrival/departure time
	Variability in actual arrival/departure time
	Number of cancellations
	Number of substitutions
	Number of diversions
	Number of operations
	Duration of GDPs

[a] We intend to examine total operational errors and deviations for flights that are impacted by a ground delay program. If there is any change in operational errors and/or deviations following GDP-E implementation at the study airport we will attempt to determine if GDP-E had any effect.

[b] Additional GDP-E performance metrics have been identified but will not be thoroughly investigated in the FFP1 operational evaluation. These metrics are discussed in Appendix B.

4.4.2 Performance Metrics

4.4.2.1 Safety

The operational outcome category, safety, has been included for all FFP1 capabilities including GDP-E. The *change in operational errors* and the *change in operational deviations* will be the primary metrics used to identify changes in system safety. The FFP1 Program Office intends to track all operational errors and deviations at many airports that experience a large number of annual GDPs. An analysis will be performed to determine if there is a statistically significant change in operational errors and/or operational deviations per arrival at each GDP location. If there is any change in operational errors and/or deviations following GDP-E implementation at the study airport we will attempt to determine if GDP-E had any effect. Since GDP-E is not intended to directly improve *or reduce* system safety, it is expected that the number of operational errors and deviations will not change with the deployment of GDP-E.

4.4.2.2 User Access

Users of the NAS experience limitations such as airspace restrictions, capacity constraints, and the number of operations that can be handled by ATC. The ability to overcome these limitations has been identified as access. One metric has been identified that is intended to capture the effects of GDP-E on user's access to the NAS and NAS resources. This metric is the *number of unused slots.*

Number of Unused Slots

The *number of unused slots* represents a useful performance metric for measuring the accessibility of flights to the GDP airport. For this metric, it is clear that a reduction in the number of unused slots represents a benefit to NAS users and an increase in user access. Further, an increase in the number of unused slots would be a disbenefit to NAS users and result in a reduction in user access. The difficulty with quantifying this metric lies in the inability to collect slot data prior to the inception of GDP-E. Without baseline slot data, we would be unable to make the correlation between changes in the number of unused slots when GDP-E is in use. Collecting the number of unused slots during the study period however might provide a useful metric for analyzing the ongoing effect of GDP-E during ground delay programs.

4.4.2.3 Delay/Efficiency

Delay in the NAS occurs for many reasons including enroute and terminal weather, airspace congestion, and mechanical breakdown. Although it is unlikely that any of the FFP1 capabilities can impact system delays resulting from mechanical breakdown, GDP-E has demonstrated the ability to positively impact the weather delays. One noted function of GDP-E is to assist in the mitigation of aggregate delay during GDPs.

GDP-E assists in reducing system delay with the introduction of better and more timely information to both users of the NAS (airlines) and the FAA (ATCSCC). This facilitates the creation and operation of more efficient ground delay programs.

Ground delay programs cause system delays by reducing available arrival demand at airports initiating the program. Generally, bad weather forces the reconfiguration of runways at an airport and the requirement that all flights use Instrument Flight Rules (IFR). These actions necessarily reduce the AAR for the GDP airport. If this reduced AAR falls below actual airport demand, flight delays and cancellations will result. The initiation of a GDP or a ground stop attempts to reduce the overall cost of these situations by holding aircraft on the ground at the source airport[4]. If a ground stop were not applied, a flight would likely be placed in an airborne holding pattern once it reached the airspace adjacent to the GDP airport until an arrival slot was available. This, consequently, would result in delay and higher cost to the airline in airborne fuel consumption.

GDP-E provides the ATCSCC with higher quality and more up-to-date information on the status of flights and the intentions of airlines. During bad weather days accurate data are necessary to guide flow management decisions. Proponents of CDM believe that this "better" information will allow the ATCSCC to more accurately predict departure times, and with more timely cancellation notices, develop more accurate demand profiles. More accurate demand profiles should in turn provide more efficient GDPs resulting in reduced delay among NAS users. In fact, it has been suggested that GDP-E has helped defer the onset of some GDPs and on more than one occasion completely prevented GDPs.

Mean Flight Time

Several metrics have been identified to capture a decrease in system delays due to GDP-E. The first, *mean flight time,* will be examined to determine if flight times have decreased during ground delay programs for certain city pairs. It has been proposed that with better information allowing the ATCSCC to create more efficient GDPs, more aircraft will be held at the source airport reducing the number of aircraft that would have to experience flight delays. As a result, the number of aircraft experiencing delay in an airborne holding pattern or the duration of delay experienced in holding patterns may decrease. If either of these operational impacts occurs, mean flight time between certain city pairs should also decrease.

Additionally, *mean flight time* may be defined as a measure of holding or delay absorption associated with arrival runway usage (*excess arrival flying time*). Using two circles with specified distances from the GDP airport, the traversal time from the outer circle

[4] For the purpose of this Metrics Plan, the source airport refers to the airport where flights are departing from on their way to the airport experiencing the grown delay program.

to the inner circle will be assessed. That traversal time will be compared with two baselines: an idealized, modeled traversal time, and a Visual Meteorological Conditions (VMC) traversal time. The VMC baseline is used to evaluate the holding or delay absorption which is the standard air traffic management practice in a certain locality. Figure 4-7 presents a graphical representation of flying time (in minutes) for SFO using 100 and 40 nmi rings.

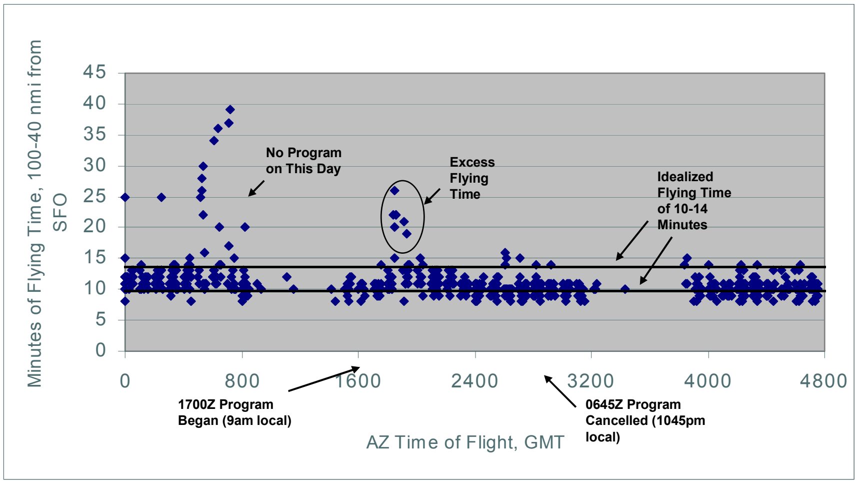


Figure 4-7. Flying Time Metric – Feb 16-17, 1999; GDP-E @ SFO

The *excess arrival flying time,* as presented in Figure 4-8, is interpreted as follows. The number of flights subjected to excess arrival flying time, and the amount of the excess might increase under GDP because of uncertainty in the true airport landing capacity and the reduced controllability of take-off times. To the degree that GDP excess arrival flying time is like the VMC baseline value, then the GDP is successful, since it keeps pressure on the runway via appropriate airborne reservoir. Conceptually, we want to ensure that holding does not increase as the ATM system seeks to fully utilize scarce runway capacity.

Compression Minutes Saved

The second metric that measures the impact of GDP-E on system delay is c*ompression minutes saved.* Compression, also known as bridging substitutions, is a process whereby unusable arrival slots are shifted in time so the owner can again use that slot. An example of how compression works is stated below.

> An airline has two flights scheduled to arrive in EWR, fight one at 1300 and flight two at 1500. After a GDP is run, flight one is assigned a 1400 arrival slot and flight two receives a 1700 arrival slot. If flight one is cancelled, flight two can't make use of the 1400 arrival slot because it occurs before its scheduled arrival time of 1500. Compression will allow the vacated slot to move down to where flight two can make use of it.

Compression savings have been collected by Metron, Inc. since September 8, 1998 for nine airports[5]. Collection began for all remaining airports that experience GDPs in October 1998. This savings is obtained using Flight Schedule Monitor (FSM) - a decision support tool that is intended to provide ATCSCC and airline specialists with on-line and historical CDM data. There are three components to FSM: a graphical and timeline presentation of demand, information extraction, and ground delay utilities. FSM permits airline participants and ATCSCC specialists to view the same information, and allows them to perform "what if" analyses and explore different traffic flow management alternatives.

It should be noted that some benefits associated with the stated compression savings could have been achieved by the airlines through their own substitutions process. Therefore, actual reported compression savings are likely high. Analysts at Metron have estimated possible airline substitutions and identified savings which the airlines alone could not achieve. This savings is identified as pure compression savings. Cumulative compression savings from January 20, 1998 to March 17, 1999 is presented in Figure 4-8.

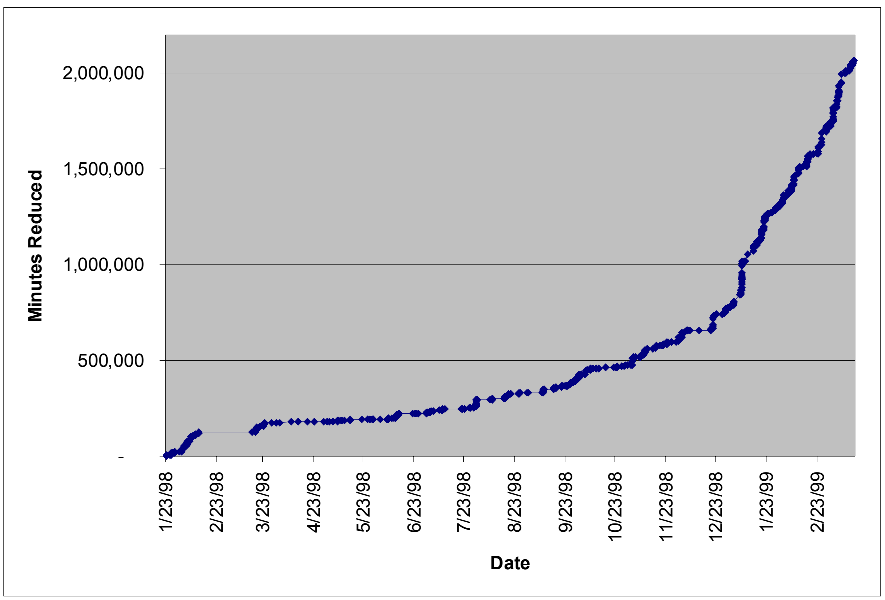


Figure 4-8. Cumulative Compression Benefits (Jan 20, 1998 – Mar 17, 1999)

Compression minutes saved is collected every time a NAS GDP is run and, using FSM, these savings may be totaled during the FFP1 evaluation schedule. The convenient aspect about tracking these savings is that FSM, the experimental testbed, has already been

[5] The nine NAS airports include ATL, BOS, EWR, LGA, MSP, ORD, PHL, SFO, and STL.

developed resulting in little effort to compile and analyze the data. In addition, *compression minutes saved* provides a good proxy to determine the actual time savings, or delay mitigation, that is achieved with GDP-E.

4.4.2.4 Predictability

As stated earlier in this document, predictability measures the variation in the ATM system as experienced by the user. There are several metrics that are intended to capture the effects of GDP-E on NAS predictability:

- Integrated Predictive Error (IPE)
- Rate Control Index (RCI)
- EDCT compliance ratio
- Number (or percent) of GDPs cancelled near start
- Number of GDP revisions

Integrated Predictive Error

In their August 1998 benefits assessment report, NEXTOR uses *the Integrated Predictive Error (IPE)* metric to assess the predictive accuracy of both flight arrival time and flight departure time estimates. The metric is designed to assess the cumulative error of a stream of predictions made over time for a single event. The NEXTOR report describes the IPE metric:

> The IPE metric assigns a single number to the stream of predictions based on how far off (in absolute value) the predictions were from the actual event. This number is called the IPE of the event. The most typical applications are measuring predictions in the departure time or arrival time of a flight. The IPE metric generalizes the "snapshot" method in which predictive errors are computed at single points in time. IPE is robust with respect to a small number of outliers in the prediction stream, provided that they are left uncorrected for short periods of time. The underlying computation in the IPE metric begins by plotting the absolute error of each prediction as a discrete function of time. This discrete function is converted into a step function by projecting each predictive error until the time of the next prediction. Finally, the step function is integrated (over time) to arrive at the IPE value.

The current analysis indicates that IPE provides a reasonable methodology in estimating the predictability of arrival and departure times. Although the results of NEXTOR's preliminary analysis did not show a significant increase in arrival time predictions between GDP-E and non-GDP-E cases, further study should be made.

Rate Control Index

The *Rate Control Index (RCI)* measures the flow of air traffic into an airport prior to any airborne holding and compares it to the targeted flow that was set by the traffic flow managers at the ATCSCC during a GDP. A single index, or percentage, is reported for the entire performance of a GDP on a single day. A higher score corresponds to better performance, meaning the flow of traffic into the airport closely matched the targeted pattern of traffic, both in quantity and in distribution. RCIs rarely go below 60% and usually hover around 90%. A perfect score of 100% is obtainable and has happened on several occasions. Since RCI does not, in and of itself, explain the good or bad execution of a program, GDPs with unusually high or low RCIs should be further analyzed for causality.

EDCT Compliance Ratio

The third predictability metric that has been identified is *EDCT compliance ratio*. Since its beginning, CDM has been providing airlines with real-time airport arrival information and has encouraged airlines to focus on EDCT compliance in a collaborative manner. The analysis will attempt to measure the possible effect of CDM on the air traffic system from an EDCT compliance perspective. Evidence suggests that, over time, CDM will increase the EDCT compliance ratio.

We compared the baseline data (data not affected by CDM generated GDPs) against the data affected by CDM generated GDPs. We need to consider flights only affected by the active GDP. This means that if the GDP is prematurely cancelled, we do not include the flights with EDCTs after the cancellation time. The period considered for this EDCT compliance study is from January 1997 to June 1999. The baseline period is January 1997 through December 1997. The period between January 1998 and August 1998 is the transition period. The CDM “in use” period is from September 1998 to June 1999. We will calculate and indicate whether the data in the transition period is close to the baseline or CDM “in use.”

Data source will include historical FSM data. The historical FSM data in 1997 was based on primarily ETMS data with some aggregate demand list (ADL) CDM data. In 1997, Volpe TCS created the CDM string, which is based on the ETMS string and includes airline data. Since 1998, the FSM data has been purely based on the CDM data. Because the EDCT compliance study is based only on the limited data filed, actual flight take off time (DZ) and EDCT (controlled departure time, CTD), the effect of the data source transition from ETMS to CDM is minimal.

This sample study considered only those flights that had an EDCT. A comparison was made between EDCT and the actual take-off time. The FAA definition of EDCT compliance was applied and the trend analysis was conducted.

Table 4-8 displays sample compliance data. It contains the month and year of the data followed by percentages of early departure, on-time departure and late departures. The definition of on-time departure is a take off between –5 and +15 minutes of EDCT. If the difference between DZ and EDCT is outside of –120 and +180 minutes, the flight is either considered invalid, or some special event took place for this particular flight and it is excluded from this analysis. This range was determined by reviewing historical data for inconsistencies.

Table 4-8. Sample EDCT Compliance Data

Month	Early	On-Time	Late
Jan-97	21.90	52.54	25.56
Feb-97	30.76	48.97	20.28
:	:	:	:
Feb-99	19.70	64.90	15.50

The analysis on EDCT compliance will include a further description data sources in addition to a trend analysis including the teams findings and conclusions. Sample data is displayed in Figure 4-9.

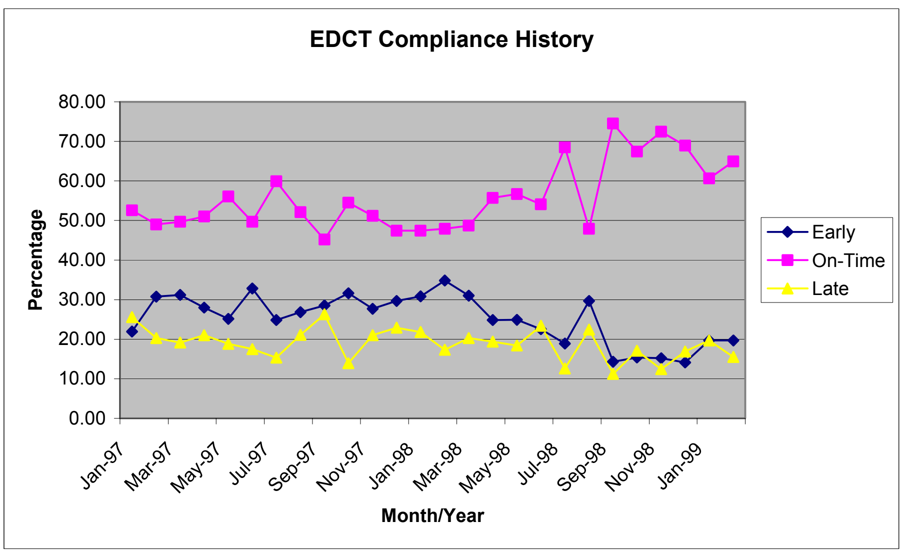


Figure 4-9. EDCT Compliance History

Number (or percent) of GDPs Cancelled Near Start

It has been suggested that the number of GDPs may decrease as a result of GDP-E. However, it will be extremely difficult to identify all of the programs that would have taken place but did not. In addition, comparing months or years prior to and post GDP-E may simply be a measure of bad weather days. Furthermore, it would be difficult to explain the reasons for the change. As a result, the *number (or percent) of cancelled GDPs* has been identified as a comparable metric.

Canceling a GDP is not often a desired action because it may indicate that delay has been unnecessarily allocated to flights. An ATCSCC specialist may cancel a GDP for reasons such as bad weather dissipating or demand not materializing. It is assumed that the inception of GDP-E allows for more predictability, flexibility, and control to plan more effective GDPs. It is not clear whether an increase in the number of GDPs represents a positive effect of GDP-E. Further, it is not clear whether an increase in the absolute differences between the planned and actual duration of GDPs indicates a positive impact of GDP-E. It was discovered that the actual duration of GDPs exceeded the planned duration by greater margins in 1998 than in 1997. Thus, a metric that provides more insight and confirmation of the benefits due to GDP-E is the percentage of GDPs cancelled.

Figure 4-9 presents the preliminary results of an analysis conducted during full prototype operations of GDP-E which began on September 8, 1998. The results shown in this histogram will be examined to determine if GDP-E had an effect on the change in the value of the metric.

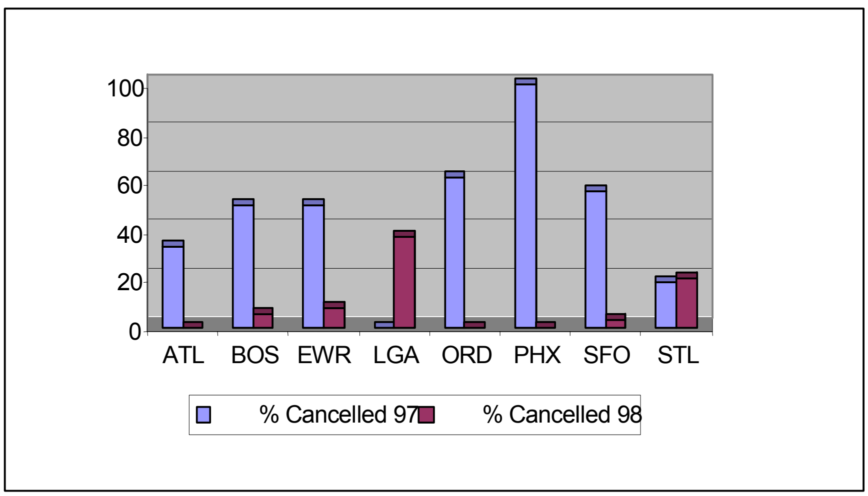


Figure 4-10. Percentage of Cancellations of GDPs from 9/8 to 12/31

Variability of Flight Distance

The number of GDP revisions will be tracked during the pre-GDP-E and post-GDP-E period. An analysis will be performed to determine if there is a statistically significant change in the number of GDP revisions at certain NAS airports pre and post GDP-E deployment. If there is any change in the number of revisions following GDP-E implementation at the study airports, we will attempt to determine if GDP-E had any effect. With better and more timely information provided by GDP-E, the number of revised GDPs may decrease.

4.4.2.5 Flexibility

Flexibility can be defined as the ability of the ATC system to meet users' changing needs in their efforts to optimize schedules, flights, and routes based on the individual objectives of the airline. Therefore, any system that aids in this optimization should contribute to increased flexibility. The FFP1 performance metrics that are intended to quantify the contribution that GDP-E makes to this operational outcome are *mean distance flown* and *Control Time of Arrival (CTA).*

Mean Distance Flown

The objectives of every airline vary by flight. It may be necessary to sacrifice the schedule of certain low-revenue flights in order to maintain the integrity of high revenue flights. Generally it must be assumed that airlines would prefer to spend less on fuel for all of their flights thereby reducing overall fuel cost. *Mean distance flown* is intended to identify changes in flight distances contributed by GDP-E. *Mean distance flown* is related to the delay metric *mean flight time* and will likely result in the same level of operational impact. Under this assumption, the additional flight distance that an aircraft must fly would take place in a holding pattern near an airport. If airport demand can be more accurately determined with the use of GDP-E one can expect the number of flights waiting in the holding pattern or the duration of time a flight spends in the holding pattern to decrease over time

Control Time of Arrival

Control Time of Arrival is a planned functionality of GDP-E. CTA will allow an airline participant, given an arrival slot at a GDP airport, to determine its own departure time according to its own objectives. This departure time would become that flight's Estimated Departure Clearance Time (EDCT). This ability will provide the airlines with more options in determining which flights they choose to be delayed. The metric used to quantify the impact of CTA is the airline's ability to comply to the give EDCT, i.e. EDCT compliance rate. CTA is currently only in an experimental version, however once the final version is completed it may be possible to examine the decision making processes that airlines would optimally choose if they have the opportunity.

4.4.3 Evaluation Schedule

The evaluation schedule for GDP-E is not dependent on the deployment of the capability since GDP-E has been in prototype operation at all GDP airports since late 1998. In short, this means that any attempt to collect baseline data, other than using automated systems such as ETMS and ASQP, will be very difficult. Baseline data will therefore be limited to data that have been archived prior to the deployment of GDP-E

Evaluating the performance of GDP-E provides us with the flexibility to tailor the GDP-E evaluation schedule according to which metric or metrics are thought to provide the best results. Based on a limited amount of data on GDPs, the evaluation plan will likely focus its data collection on those airports that incur a large number of annual GDPs. This will provide the largest amount of data to identify significance with any of the chosen metrics. Table 4-9 identifies the top six GDP airports and their average annual number of GDPs they have experience over the last four years.

Table 4-9. Top GDP Airports and Average Number of GDPs(1995-1998)

GDP Airport	Annual GDPs[6]
San Francisco (SFO)	163
St. Louis (STL)	70
Newark (EWR)	67
Boston (BOS)	58
Chicago (ORD)	57
Philadelphia (PHL)	24

4.4.4 Data Collection

An evaluation plan of this magnitude requires the collection and storage of a large amount of data. This will not be difficult for those metrics that require collection of ASQP or CODAS data since the data is being archived by ASD-400 and APO. However, for the metrics that will be analyzed using ETMS or ARTS data, an extensive database will need to be developed.

The NEXTOR preliminary report will further define the GDP-E performance metrics and the measurement methodology used to quantify each metric.

[6] These numbers represent the average annual number of ground delay programs that are experienced by the airport between 1995 and 1998.

4.4.5 Evaluation Issues/Concerns

Delay/Efficiency

Compression minutes saved. The compression algorithm calculating minutes saved may include delay savings that could have been captured through an airline's internal substitution process. However, some airlines do not have the capability to do internal substitutions, or during some GDPs do not have the opportunity to make these changes within the compression window. Additionally, the compression algorithm does not capture minutes saved where GDPs are avoided completely or are short-ended in duration. Currently, analysts at Metron Inc. are developing an alternative methodology for calculating compression minutes saved. This new methodology promises to report a more conservative amount of savings.

Predictability

Number of GDPs cancelled near start. The number of GDPs cancelled presents several challenges to explain the benefit of the change. Cancelled GDPs may be either "good" or "bad" depending on when they take place. A GDP cancelled prior to the start of the program may be considered "bad" since delay has already been experienced by the airlines. In this situation, the cancellation of the GDP led to delay despite the GDP being cancelled.

If a GDP is cancelled after the GDP has been in effect for some time it may considered a good impact. At this point airlines are expecting a major impact in delay to existing flights and the possible cancellations of other flights. The sooner the GDP is cancelled, the sooner the airlines will be able to recover from any delay that is impacting their operations.

For those "bad" cancels that occur before the start of the program it may be necessary to examine the number of characteristic of the ground stop that succeeded or preceded the GDP. This may allow some insight as to whether the cancelled GDP had either a positive or negative overall impact.

4.5 CDM: NAS Status Information

4.5.1 Evaluation Overview

NASSI is a key component of CDM that provides for the sharing of data about the operational status of the NAS.

The availability of NASSI is expected to increase airline understanding of ATM intentions and actions. A better understanding of ATM motives for calling a particular action will result in a decrease in user workload; the airlines would not have to expend energy guessing the next ATM action. It is also expected that the amount of time currently dedicated to negotiations between AOCs and ATCSCC will be reduced. This time savings for both users and ATM service providers is expected to increase NAS flexibility and provide user's access to services on demand. This concept of a common view of the same data by all NAS participants promotes a shared understanding of the decisions that must be made to manage NAS traffic – this concept of greater understanding is a accepted direct benefit to all NAS users.

Note that NASSI is still in the earliest stages of definition for deployment. Currently, FSM-related data items are at the stage of IDU, i.e. AAR, GDP projected demand and capacity, planned pushback times. The data sources of the remaining elements is still being defined. The number of sites and specific data items for each of the sites is still being defined. Therefore the operational impacts and performance metrics outlined above are subject to change once the final NASSI scope and functionality are defined.

4.5.2 Performance Metrics

No analyses are planned, however the Metrics Team will continue to survey the NASSI users on their perception of the operational impact of the capability. The CDM Product Team will also collect usage data based on the number of "hits" on the NASSI website.

4.6 CDM: Collaborative Routing

Collaborative Routing (CR) is defined as information sharing for creating and assessing rerouting strategies around hazardous weather, SUAs, and other constrained airspace. This capability employs electronic "chalkboards" for use by ATCSCC, en-route Center TMCs, and ATC coordinators at AOCs for display conferencing. The application of this collaborative "display conferencing" capability results in greater common situational awareness, faster decision making, and common understanding of solutions. This in turn means greater flexibility for the airspace users in their flight planning and defining their desired route. The user may also observe overall flight efficiency, possibly resulting in a reduction in delay, as all participants have one common view of the constrained airspace and thus improves the quality of their decisions.

4.6.1 Evaluation Overview

Collaborative Routing seeks to provide benefits through the collaboration and real-time information sharing among airspace managers and airspace users. The evaluation of such a capability will most likely rely on qualitative measures and subjective data from the CR users. The CR performance metrics are an attempt to assess the operational impact of CR on NAS operations in a quantitative manner, but it must be realized that this type of evaluation may not result in any measurable impact on NAS operations.

FFP1 are currently deploying the technologies to seven sites to support the "display conferencing" capability (PicTel data conferencing over ISDN lines). Procedures are being developed and field trials are being conducted to evaluate potential benefits of this CR function. Additional components that support and contribute to the success of CR are also being evaluated. These components include prototype CR coordination functions, consensus convective weather forecasts, coded SWAP routes, and other procedures. Consideration for further deployment of the "display conferencing" capability, e.g. AOCs, is also being evaluated.

The CR functionality is under evaluation by the FFP1 PO in cooperation with Stakeholders. Therefore the operational impacts and performance metrics are subject to change. Further analysis is needed to define the evaluation methodology used to quantify the expected impacts of CR on NAS users and service providers.

4.6.2 Performance Metrics

The CR Performance Metrics are summarized in Table 4-10. These performance metrics were defined by assessing the expected operational impact of CR and identifying how this impact works to achieve FAA's operational outcomes. No analysis has been performed to further define these metrics at the time of publishing this Metrics Plan. In future reports on the implementation of this Metrics Plan, the CR capability and its operational impacts on the NAS users will refer to a NEXTOR ICR Evaluation Plan.

Table 4-10. CR Performance Metrics

Operational Outcome	Performance Metric
Safety	Change in operational errors while capability is in use
	Change in operational deviations while capability is in use
User Access	Number of operations
	Number of aircraft using Special Use Airspace (SUA)
	Number of diversions
Delay/Efficiency	Average flying time
	Standard deviation of predicted fuel usage and actual fuel usage
Flexibility	Number of user preferred routes flown
	Average flying distance

4.7 Surface Movement Advisor

SMA was originally developed by NASA as a tool to aid ramp and FAA tower controllers to more effectively plan ramp movements and coordinate ground support operations. The prototype SMA system was installed at Atlanta Hartsfield International Airport (ATL) and was found to significantly decrease taxi-out times there. Since that time SMA has evolved from a tool intended to assist in ramp management to a more general "information sharing" concept. Under FFP1, SMA consists of a data feed from the ARTS radar processor at the airport where the tool is implemented to any airport users who wish to obtain this data.[7] A computer application has also been developed, which is provided free of charge to potential SMA users, that displays radar tracks of arriving and departing aircraft at the SMA airport, along with aircraft identifications and estimated times of arrival (for arriving traffic) at the airport. This data and the associated computer application may be used by airlines at their Airline Operations Centers (AOCs) or their airport stations, by airport ramp controllers, or by other ground service providers. Users may even incorporate the SMA data into their own computer applications.[8] The potential impact of SMA deployment at AOCs needs to be tracked for any influence on FFP1 metrics for SMA.

4.7.1 Evaluation Overview

Since SMA will be used by airlines, ramp controllers, airport operators, and ground service providers to help manage operations at an airport or in the terminal airspace environment, our evaluation of SMA performance focuses on the contribution of the system to FAA performance outcomes within the terminal airspace and at the airport surface. The evaluation will examine changes in taxi times, gate delays, gate reassignments, and diversions at the SMA airports attributable to SMA usage. Additionally, any changes in safety (as indicated by operational errors) will be analyzed and reported. Where possible, data for each metric will be collected for a period of time prior to SMA implementation in order to establish baseline performance, and then for a period of time after SMA implementation so that the effect of SMA may be assessed. Local environmental, airport configuration, and airport demand data will also be collected in order to assess how these factors affect SMA usage and benefits, and in order to isolate the effects of SMA from those of changes in these conditions. Data will be collected for a period of one year prior to system deployment and one year following system deployment at each site so that seasonal

---

[7] The FFP1 Program Office distinguishes between the "Atlanta SMA" developed by NASA and the "FFP1 SMA" that will be fielded to the FFP1 CCLD sites.

[8] For example, Northwest Airlines may incorporate the SMA data into the Datalink Delivery of Expected Taxi Clearance (DDTC) system that they use at Detroit Wayne County Airport.

factors may be fully removed, and so that any learning curve associated with SMA may be observed.

The SMA performance metrics are summarized in Table 4-11, and each is described in the following sections. For each metric we discuss the reason why the metric was selected, the direction in which we expect the metric to change following implementation of the particular tool, the data that will be required to calculate and understand the metric, and the data sources that we expect to rely on. At the end of this section we describe any concerns or issues related to computation or assessment of these metrics. All metrics will be computed for peak periods, where a peak is defined as a period where the actual arrival rate is close to the airport acceptance rate, not necessarily a high actual arrival rate or high airport acceptance rate. The metrics are organized by the FAA Operational Outcomes to which they relate. Note that the metrics associated with taxi times and gate delay were developed based on airport usage of SMA (ramp tower).

Table 4-11. SMA Performance Metrics

Outcome Category	Metric
Safety	Change in operational errors when capability is in use
	Change in operational deviations when capability is in use
User Access	Diversion rate
Delay/Efficiency	Mean taxi-in time
	Mean taxi-out time
	Mean gate delay
Predictability	Variability of taxi-in time
	Variability of taxi-out time
	Variability of gate delay
	Gate reassignment rate

4.7.2 Performance Metrics

4.7.2.1 Safety

The FFP1 metrics that are intended to address system safety are the *change in operational errors* and the *change in operational deviations*. The FFP1 Program Office intends to track all operational errors and deviations that occur at each SMA site. An analysis will be performed to determine if there is a statistically significant increase in operational errors or deviations per operation at each site after the implementation of SMA. Since SMA is not intended to directly improve or reduce system safety, it is expected that the number of operational errors/deviations per operation will not change following SMA implementation, and it is desired that the number of operational errors/deviations per operation will actually decrease. In addition to examining the overall number of operational errors/deviations and the number of operational errors/deviations per operation, the cause of each operational error/deviation will be identified, and those that are either a direct or indirect consequence of SMA usage will be reported.

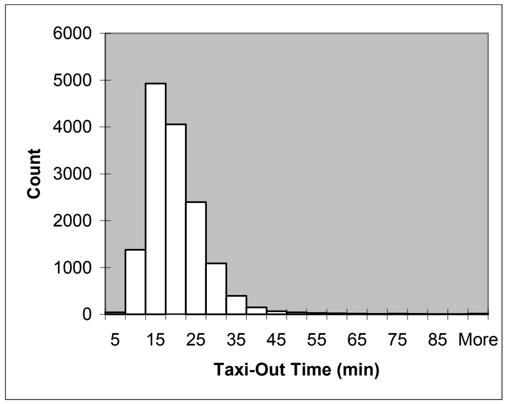
4.7.2.2 User Access

The metric that captures SMA's contribution to increasing users' access to the NAS is the *diversion rate*. Northwest Airlines, which is currently using an SMA system in their operations center to help manage arrivals at Detroit Wayne County Metropolitan Airport, reports that they are able to avoid about two diversions per week at Detroit because of the enhanced situational awareness made possible by SMA. We intend to measure the *diversion rate* (defined as the ratio of diverted flights to the total number of arrivals for a given time period) at airports such as Detroit where SMA is being used by an airline's operations center. The goal of the metric is to capture avoided diversions. We will examine this rate both before and after SMA implementation, and determine whether there has been a statistically significant change in diversions. Weather data will also be closely examined, since severe weather can have a profound impact on the number of flights diverted. In addition to the proposed quantitative analysis of diversions, we intend to hold discussions with airlines utilizing SMA in order to obtain their perceptions of changes in diversion rates attributable to SMA (the variability of diversion rates may not allow a quantitative differentiation between the pre- and post-SMA time periods).

4.7.2.3 Delay/Efficiency

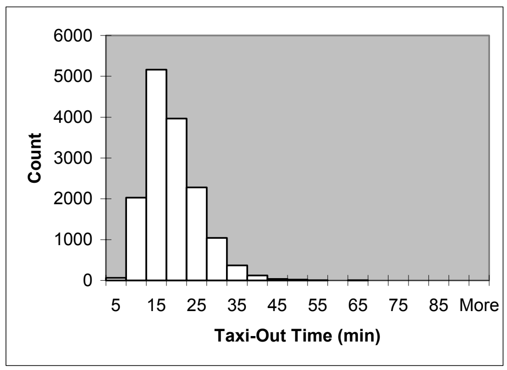
The primary metrics intended to capture SMA's contribution to NAS delay reduction are the *mean taxi-in and taxi-out times*. At airports where SMA is installed in the ramp tower, taxi times are expected to significantly decrease. Several studies have found that taxi-out times decreased by approximately one minute per operation at ATL after SMA was installed in the ramp towers there. Figure 4-11 illustrates the change in distribution of taxi-out times observed at ATL.

Baseline



Mean	18.13
Median	17
Mode	15
Sample Variance	66.33
Minimum	1
Maximum	150
Count	14652
Confidence Interval(95%)	0.131888

SMA



Mean	17.03
Median	16
Mode	13
Sample Variance	44.48
Minimum	1
Maximum	101
Count	15098
Confidence Interval (95%)	0.106

Figure 4-11. SMA Taxi-Out Time Comparison, Atlanta Hartsfield International

The other SMA delay metric selected is *mean gate delay*. Gate delay is defined as the difference between the published scheduled departure time and the actual push-back time. Further analysis has shown that in addition to reducing taxi-out times at ATL, SMA also resulting in a decrease in gate delays of approximately four minutes per operation.

The contextual data that will be collected along with these delay metrics include surface weather (e.g., ceiling, visibility, and precipitation rate), arrival rate, and departure rate. In addition, if gate and runway assignment can be obtained for each operation a more detailed analysis of taxi-times may be possible than if simply airline, flight number, and equipment are known.

4.7.2.4 Predictability

Three of the four SMA predictability metrics reflect the variability of the three delay metrics discussed above:

- *Variability of taxi-in time*
- *Variability of taxi-out time*

- *Variability of gate delay.*

If SMA can reduce the variability of taxi times and gate delays, airline operations will be more predictable, and hence airlines will be able to reduce costs. As Figure 4-11 indicates, there was a considerable reduction in the variance of taxi-out time at ATL once SMA was implemented.

Variability of these three measures will be reported using the standard deviation, e.g.,

$$V_{GateDelay} = \sqrt{\frac{\sum_{i=1}^{n}\left[\left(t_{SchedDepart_i} - t_{ActDepart_i}\right) - \frac{1}{n}\left(\sum_{i=1}^{n}\left(t_{SchedDepart_i} - t_{ActDepart_i}\right)\right)\right]^2}{n-1}}$$

The contextual data that will be collected along with these delay metrics include surface weather (e.g., ceiling, visibility, and precipitation rate), arrival rate, and departure rate. In addition, if gate and runway assignment can be obtained for each operation, a more detailed analysis of taxi-times may be possible than if simply airline, flight number, and equipment are known.

The other predictability metric for SMA is the *gate reassignment rate*. Delta Airlines has reported that the improved ramp management resulting from SMA usage at ATL has led to a 30 to 40 percent reduction in gate reassignments at the airport (a gate reassignment occurs when an arriving aircraft is sent to a gate other than that originally planned, which requires the costly repositioning of ground resources). The *gate reassignment rate* will be computed as the ratio of the number-of-gate-reassignments to the number-of-arrivals for a given period of time.

4.7.3 Evaluation Schedule

The SMA evaluation schedule is illustrated in Figure 4-12. Baseline data will be collected for a period of one year prior to deployment (PCA) at each site where possible. In-use data will be collected for a period of one year following deployment. Data will not be analyzed from ATL, since the system there has already been extensively analyzed, and the FFP1 SMA implementation is considerably different from the NASA implementation.

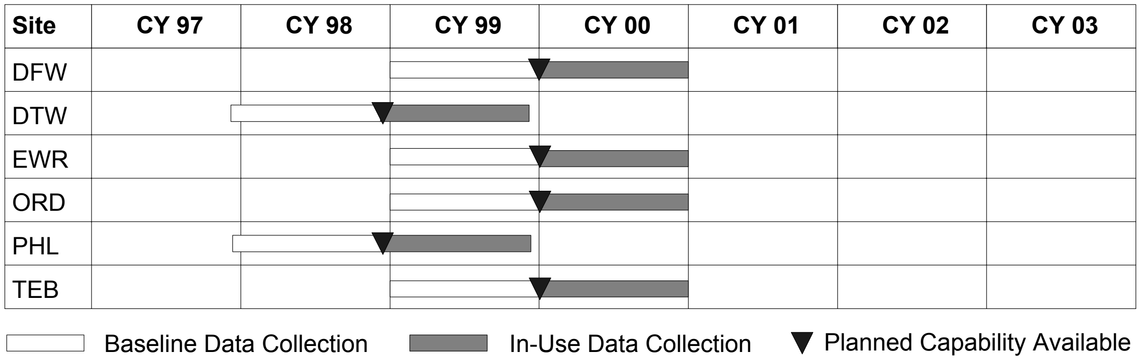


Figure 4-12. SMA Evaluation Schedule

4.7.4 Data Collection

The SMA metrics and associated data elements are outlined in Table 4-12, along with the sources for these data elements (for those data elements to be collected by the FAA) and the frequency with which they will need to be collected. Each metric is discussed in more detail below (note that some of the metrics have been grouped together as they rely on the same data sources and will be evaluated using the same context data).

Number of operational errors/deviations attributable to SMA - these metrics will be derived from the National Aviation Safety Data Analysis Center (NASDAC) database.

Mean/variability of taxi-in/taxi-out time - these metrics will be derived from ASQP data. Arrival and departure rates will be derived from ETMS data or from the SMA ARTS data once the systems have been implemented. Gate and runway assignment data will need to be provided by the airlines. Hourly surface weather data are collected and distributed by the National Climatic Data Center.

Mean/variability of gate delay - this metric will be derived from ASQP (actual departure time) and OAG (scheduled departure time) databases. Actual arrival and departure rates will be collected from either ETMS or the SMA ARTS data feed. Hourly surface weather observations are available from the NCDC.

Gate reassignment rate - The planned and actual gate assignments for each flight at SMA airports will need to be provided by the airlines. Arrival and departure rate information may be obtained from ETMS or from the SMA data feed. Hourly surface weather observations are available from the NCDC.

Diversion rate - the number of diversions at SMA airports where AOCs use the tool will be derived from ASQP data. Associated flight data may also be obtained from ASQP. The actual arrival rate may be obtained from ETMS or the SMA data feed. Surface weather data may be obtained from the NCDC.

Table 4-12. SMA Metrics, Data Elements, and Data Sources

Outcome Category	Metric	Data Elements, *Local Factors*	Frequency	Data Source
Safety	Change in operational errors while capability is in use	Operational error report	on occurrence	NASDAC, Facility records
	Change in operational deviations while capability is in use	Operational deviation report	on occurrence	NASDAC, Facility records
Delay/ Efficiency, Predictability	Gate reassignment rate	Number of gate reassignments	daily	Calculated
		Arrival rate	30 minutes	SMA ARTS data feed
		Airport	per operation	ASQP
		Call sign	per operation	ASQP
		Actual gate	per operation	Airlines
		Planned gate	per operation	Airlines
		Departure rate	30 minutes	SMA ARTS data feed
		Ceiling, visibility, precipitation rate	hourly	NCDC
	Mean taxi-in time, Mean taxi-out time, Taxi-in time variability, Taxi-out time variability	Taxi time	per operation	ASQP
		Airport	per operation	ASQP
		Call sign	per operation	ASQP
		Actual arrival/departure time	per operation	ASQP
		Equipment type	per operation	SMA ARTS data feed
		Gate	per operation	Airlines
		Runway	per operation	Airlines
		Arrival rate	30 minutes	SMA ARTS data feed
		Departure rate	30 minutes	SMA ARTS data feed
		Taxi distance	once	Site survey
		Ceiling, visibility, precipitation rate	hourly	NCDC
	Mean gate delay, Gate delay variability	Gate delay	per operation	calculated
		Airport	per operation	ASQP
		Call sign	per operation	ASQP
		Actual departure time	per operation	ASQP
		OAG departure time	per operation	OAG
		Departure rate	30 minutes	ARTS or TRACON logs
		Arrival rate	30 minutes	ARTS or TRACON logs
		Ceiling, visibility, precipitation rate	hourly	NCDC
	Diversion rate	Number of diversions	per diversion	ASQP
		Arrival rate	30 mintues	SMA ARTS data feed
		Airport	per diversion	ASQP
		Call sign	per diversion	ASQP
		Ceiling, visibility, precipitation rate	hourly	NCDC

4.7.5 Evaluation Issues/Concerns

Additional benefits metrics will need to be developed for Teterboro Airport since there is no scheduled airline service at this airport. There is, however, a great deal of General Aviation (GA) activity at Teterboro. It is expected that SMA will be used by Fixed Base Operators (FBOs) or other ground service providers. Metrics that capture the benefits to these service providers will be developed in the future.

Safety

The metrics used here to capture safety effects of SMA are the *number of operational errors* and the *number of operational deviations*. SMA is not generally used by air traffic controllers or FAA employees, but by airline operations personnel and ground service providers. These metrics will therefore only have meaning at those SMA sites where a system is installed in the Air Traffic Control Tower (ATCT). There currently is no plan to install SMA in any ATCTs. Even if systems were installed in ATCTs, the operational error/deviation rate in towers is so small that it would be extremely unlikely that one could attribute any change in the rate to SMA.[9]

User Access

Diversion rate will only be considered at those SMA airports where a display is installed and used at an AOC to manage arrivals, such as DTW. Furthermore, weather must be carefully considered when analyzing the changes in the number of flight diversions, since severe weather will have a significant impact on this metric and could easily mask any impact that might be produced by SMA.

Delay/Efficiency

SMA displays may not actually be installed or used at some of the SMA airports. For example, an SMA data feed from the ARTS radar processor at Detroit's Wayne County Metropolitan Airport was implemented in December 1998, but the display was only installed at the Northwest Airlines Systems Operations Center (SOC) in Minneapolis, Minnesota. We do not expect to see any reductions in taxi times at DTW resulting from SMA implementation, since the system is not being used to manage surface operations. Therefore *mean taxi-in/taxi-out time* will not be considered at SMA sites where the system is not installed at ramp towers or ATCTs.

For those sites where SMA is utilized at the airport, ASQP data will be relied on to compute taxi-in and taxi-out times. It should be noted, however, that ASQP has a precision of only one minute, and the accuracy of the data varies by airline and crew. Furthermore, there are no ASQP data for commuters.

Predictability

Variability of taxi-in/taxi-out time are subject to the same ASQP data problems as are the mean taxi times discussed above.

[9] The operational error rate in ATCTs has historically been about 0.23 operational errors per 100,000 operations.

Gate reassignment rate will only be considered at those SMA sites for which a display is actually installed and used at the airport for ramp management purposes.

Section 5

Evaluation Roles and Responsibilities

5.1 Free Flight Phase One Metrics Team

The role of the FFP1 Metrics Team is described in this Metrics Plan. In short, the Metrics Team will lead the data collection, analysis, and reporting on the FFP1 operational impacts. The Metrics Team will interface with users through periodic meetings to insure continued collaboration on the performance assessment of FFP1 capabilities. FFP1 Metrics Team is part of FFP1's cross-cutting integration management team. In this capacity, the Metrics Team contributes to program-level decision making, including risk management, human factors, procedures, and resource allocation tradeoffs.

5.2 FFP1 Users and Service Providers

To date, the airlines (FFP1 users) have expressed their intent to support the deployment and evaluation of the FFP1 capabilities. They have expressed interest in access to the data that would be made available to them, in exchange for sharing some of their information. While the specific data elements to be shared between airlines and the FAA have yet to be defined, there is every indication that the airlines will participate as needed to support the implementation and evaluation of the FFP1 capabilities. Their participation will include providing the Metrics Team with data describing fuel usage, pilot/dispatcher intent, and final benefit values, in terms of dollars, for the results of the metrics presented in this Metrics Plan. The anticipation of industry-wide cooperation is a precedent-setting aspect of the arrangements for FFP1 deployment.

The Stakeholder groups also represent the service providers of these FFP1 capabilities. The Metrics Team acknowledges the need for cooperation from these individuals in assessing the impact of these capabilities on their workload. They will also serve a major role in assessing the impacts of these capabilities on NAS safety. The process of data exchange between the Metrics Team and the service providers has not yet been established.

List of References

1. RTCA. *FFP1 Core Performance Metrics*, December 10, 1998.
2. RTCA. *Government/Industry Operational Concept for the Evolution of Free Flight—Addendum 1: Free Flight Phase 1 Limited Deployment of Select Capabilities (URET, TMA (SC), pFAST, CPDLC, CDM, SMA)*, August 20, 1998
3. FFP1 Program Office. *FFP1 Program Master Plan: Update,* January 31, 1999.
4. Crown Communications Inc. CTASDS-BAPRPT-002. *Center/TRACON Automation System Passive Final Approach Spacing Tool (pFAST) Assessment – Final Report*, December 5, 1996.
5. Carr, Gregory C., Heinz Erzberger, and Frank Neuman. *Airline Arrival Prioritization in Sequencing and Scheduling*, 2nd USA/Europe Air Traffic Management Research and Development Seminar, Orlando, FL. December 1-4, 1998.
6. Zelenka, Richard, Roger Beatty, and Shawn Engelland. *Preliminary Results of the Impact of CTAS Information on Airline Operational Control*. AIAA-98-4481, 1998.
7. Rada, Willma, MCA Research Corp. *Surface Movement Advisor (SMA) Benefit Analysis*, October 14, 1997.
8. Post, Joseph and Jeffrey Biros, FAA ASD/SETA. *Surface Movement Advisor ATL Performance Assessment*, DCN B80205-01, August 20, 1998.
9. NEXTOR (Ball, Michael, Hoffman, Robert), *Collaborative Decision Making in Air Traffic Management: A Preliminary Assessment*, The National Center of Excellence for Aviation Operations Research (NEXTOR), DRAFT: August 1998.
10. Metron Inc. *Graphical Illustration of the CDMNet*, http://www.metsci.com/cdm/cdmnet.html, March 1999.
11. Metron Inc. CDM Elements, http://www.metsci.com/cdm/cdmnet.html, March 1999.
12. Howard, Ken and Gilbo Eugene, John A. Volpe National Transportation Systems Center. *Predictive Data Quality; Cancels, Substitutions, and Exchanges in GDP-E – Memorandum*, March 1999.
13. DeArmon, Jim, MITRE/CAASD. *Metrics for Evaluation of GDP-E Benefits: Assessment of Flying Time & Distance, Fuel Usage*, March 1999.
14. Inniss, Tasha, University of Maryland. *Percentage of Canceled Ground Delay Programs (GDPs)*, March 1999.
15. Hoffman, Bob, University of Maryland. *Rate Control Index*, March 1999.

16. Tanino, Midori, Metron Inc. *Number of Ground Stops; Number of Diversions; Number of Unused Slots; EDCT Compliance*, March 1999.

17. Brennan, Mike, Metron Inc. Planned vs. Actual Departure and Arrival Times – Average and Variance, March 1999.

18. FAA. *Air Traffic Quality Assurance (7210.56A)*, May 1, 1999.

Appendix A

RTCA FFP1 Core Performance Metrics

RTCA Free Flight Phase One Select Committee: FFP1 Core Performance Metrics

Overview

This paper briefly summarizes the purpose, organization, usage, and future direction of the Free Flight Phase 1 (FFP1) Core Performance Metrics.

Purpose

The FFP1 Core Performance Metrics were developed under the auspices of the RTCA Free Flight Steering Committee by a consensus of aviation Stakeholders from government, industry, and the user community. The purpose of the FFP1 Core Performance Metrics is to guide the collection and analysis of data to support the evaluation of FFP1 operational impacts as the FFP1 Core Capabilities Limited Deployment (FFP1 CCLD) program proceeds at multiple sites.

Organization of the FFP1 Core Metrics

The FFP1 Core Performance Metrics are summarized in the attached table. Each metric can be traced to one of six Operational Outcomes (adapted from the FAA Air Traffic Services' operational performance outcomes of the *ATS Performance Plan Fiscal Years 1998-2000*) listed in the left-hand column of the table. The Operational Outcomes are:

- Increase System Safety
- Decrease System Delays
- Increase System Flexibility
- Increase System Predictability
- Increase User Access
- Increase System Productivity

The second column of the table, Aggregate Metrics, lists the high-level metrics associated with each Operational Outcome. The remaining columns list the specific metrics associated with each FFP1 Core Capability. These capability-specific metrics are traceable both to the Operational Outcomes and to the Aggregate Metrics.

Additionally, the metrics for each FFP1 Core Capability are aligned with the four FFP1 Operational Concepts listed across the top of the table. The FFP1 Operational Concepts, described in detail in the RTCA document *Government/Industry Operational Concept for the Evolution of Free Flight, Addendum 1: Free Flight Phase 1 Limited Deployment of Select Capabilities* , are:

- Collaborative Decision Making
- Enhanced Information Sharing
- Automated Decision Support Tools
- Enhanced Communications

Using the Metrics

The FFP1 Core Performance Metrics will be used to support decisions regarding future NAS-wide implementation of FFP1 capabilities. Specifically, the metrics will help to quantify the operational impacts of the FFP1 CCLD and enable estimates of potential NAS-wide benefits. It is recognized that not all of the benefits of the FFP1 capabilities will be realized within the short timeframe of the limited deployment. Achievement of full benefits will occur over a longer period of time as capabilities and procedures mature and become more closely integrated. Nevertheless, objective analysis of the core performance metrics will provide the basis for estimating the full range of benefits enabled by the FFP1 capabilities.

Future Direction

FFP1 CCLD will be implemented during the 1999-2002 period at the sites listed in the August 1998 letter from the RTCA Free Flight Steering Committee. Given the near-term deployment schedule, collection of National Airspace System (NAS) performance data must begin immediately in order to establish a baseline upon which to measure the operational impacts of FFP1 CCLD.

As baseline data collection proceeds, additional data collection to support the FFP1 Core Performance Metrics will commence with the initial fielding of the first Core Capability and continue for multiple capabilities as implementation continues. It is anticipated that there will be synergistic operational impacts resulting from the combined effects of multiple capabilities. Also, the collateral effects of the capabilities will be measured at satellite airports and facilities surrounding the FFP1 CCLD sites.

Data collection will continue beyond the time of the last site fielding of the final Core Capability to allow for operations to stabilize as capabilities and procedures become further integrated.

Analysis of Core Performance Metrics data and assessment of FFP1 CCLD operational impacts will focus on three interdependent performance levels:

- NAS Level
- Site Level
- Core Capability Level

The three performance levels are depicted in Figure 1.

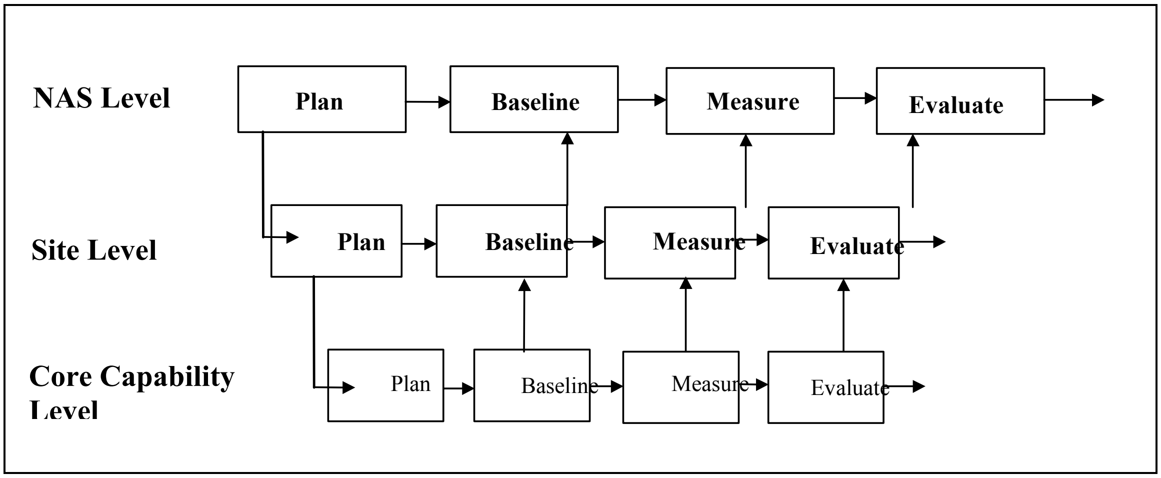


Figure 1. FFP1 CCLD Performance Measurement Hierarchy

The FFP1 CCLD Performance Measurement Hierarchy depicted in Figure 1 represents a top-down planning/bottom-up measurement and evaluation philosophy. The premise of this philosophy is that planning at the higher levels should drive planning, data collection, and performance evaluation at the subordinate levels. Subsequently, baselining, measurement, and evaluation results at the lower levels should support higher-level activities.

To initiate the planning process, the RTCA Free Flight Steering Committee will oversee the development of a FFP1 Operational Impact Evaluation Plan during the first quarter of calendar year 1999. This plan will supplement the RTCA FFP1 Core Performance Metrics and provide guidance to the FFP1 Program Office and the other stakeholders regarding the coordination of FFP1 Core Capability performance measurements at each of the three levels depicted in Figure 1.

Table A-1. RTCA FFP1 Core Performance Metrics

FFP1 Operational Concepts →		**Collaborative Decision Making**		**Enhanced Information Sharing**	**Automated De**	
Operational Outcomes	**Aggregate Metrics**	**GDP-E**	**Collaborative Routing**	**SMA**	**pFAST**	
Increase System Safety	Number of operational errors attributable to each tool *					
Decrease System Delays	Average flying time	Average flying time; Number of GDP-Es; Duration of GDP-E	Average flying time		Average difference of time cross meter fix to time cross threshold	Average cross 20 to time c
	Average difference of planned time versus actual time (arrival time, departure time)	Average difference of planned time versus actual time (arrival time, departure time)				Average planned versus a arrival
	Average taxi time	Average taxi-out time		Average taxi times; Average gate delay		
Increase System Flexibility	Lateral deviations flown		Number of user preferred routes flown			
	Average flying distance	Average flying distance	Average flying distance		Average flying distance from meter fix to threshold	Average from 20 meter fi
	Number of restrictions eliminated					
Increase System Predictability	Variability in scheduled times versus actual times	Variability in planned versus actual arrival time; Variability in planned versus actual departure time		Variability in taxi times; Variability between scheduled and actual pushback times, i.e. gate delay		Variabil expected at 200 n at meter
	Variability in acceptance rates versus actuals	Variability between arrival rates (planned and actual)			Variability between average arrival rate and acceptance rate	Variatio Variabil acceptan through
	Variability in flying times and distances					Variabil times/di nmi ran
	Average fuel usage (predicted versus actual)	Average fuel usage	Average fuel usage		Average fuel usage	Average
Increase User Access	Average throughput	Number of operations; Number of unused slots; Number of cancellations, and substitutions	Number of operations; Number of diversions; Number of aircraft using SUA		Actual arrival rate on each runway	Actual a
Increase System Productivity	Average throughput per sector or position				Distribution of aircraft on runways and sectors	Average sector o

* Subject to final review and analysis by stakeholders.

Appendix B

Changes in Metrics from Previous Versions of the Metrics Plan

This Appendix is included to document those metrics that are no longer recommended to evaluate the operational impacts of the FFP1 capabilities. In some cases, the metrics were duplicates of other metrics. In other instances, the Metrics Team has conducted initial analyses on the usefulness of the metric, and has redefined or refined the metric to capture the true essence of the user-based operational outcome. The results and findings are presented below.

B.1 User Request Evaluation Tool

The metric defined as *time spent at or near desired route for city pairs* measures how close to their desired lateral path the NAS permits users to fly. Another way of interpreting this metric is to measure the lateral deviation between the flight planned route and route flown. ATC restrictions frequently preclude users from filing the routes they desire. In these situations, pilots may request route amendments, and so the flight planned route may not reflect what the user wanted. However, flights operated according to NRP are not restricted to ATC-preferred routes and pilots frequently do not request route amendments. Therefore, this metric should be used for NRP flights only. It is acknowledged that NRP flights have already received a benefit, analysis of this metric using before and after URET data would show if URET provides an additional benefit.

This metric is no longer recommended because its intent is duplicated by other metrics (excess distance and number of restrictions eliminated) which are more direct measures of user benefits.

B.2 Traffic Management Advisor – Single Center

In this version of the Plan we have added the safety metric of *number of operational deviations*. Previously, our only safety metric was the *number of operational errors*. In general, an operational error occurs when the applicable separation minima between two or more aircraft are violated as a result of some failure of the air traffic system. In an operational deviation, separation minima between aircraft are not violated, but aircraft penetrate airspace or encroach upon a landing area delegated to another position or facility without prior coordination. We feel that the number of operational deviations can provide important additional insights into the safety of the NAS. Although an operational deviation does not necessarily indicate the occurrence of an unsafe situation, it does suggest that there has been a failure of communication between positions or facilities, and this could be the precursor to an unsafe situation. Alternatively, procedures may need to be changed if certain operational deviations repeatedly occur without any safety implications.

We have removed the metric *mean flight distance from the 200 nmi range ring to the meter fix* from this version of the Plan. This metric appeared in previous versions and in the RTCA's Core Metrics. We have concluded that flying distance is not a suitable metric for flexibility for tools that operate in the extended terminal area. By itself flying distance does not tell us anything about the operator's intent, so we would not be able to use it to measure flexibility. Flying distance is related to fuel usage, but this is specifically addressed by the delay/efficiency metrics.

We have reclassified the fuel usage metrics to the delay/efficiency category from the predictability category, as we feel that this is more appropriate. We also have modifed these metrics to focus on fuel usage from the 200 nmi range ring to the meter fix rather than simply gate-to-gate fuel burn.

We have changed the previous metric of *variability of acceptance rate less actual arrival rate* to *mean difference of airport acceptance rate less actual arrival rate*. We feel that the mean difference between the actual rate and the AAR is a simpler and better indication of predictability of an airport's throughput than the variability of the difference between the actual rate and the AAR.

We have removed the *variability of distance flown from the 200 nmi range ring to the meter fix* metric since we feel that distance flown is not a good measure for predictability in the extended terminal area (we still will measure the variability of flight times in this area).

We have slightly changed the wording of the productivity metric from *mean arrival rate per sector or position* to *mean actual arrival rate/throughput per sector or position*. We intend to calculate both the mean arrival rate per position as well as the mean throughput for individual sectors.

B.3 Passive Final Approach Spacing Tool

We have added the safety metric of *number of operational deviations*, as described previously for TMA.

We have removed the flexibility metric *mean flight distance from the meter fix to the runway threshold*, which appeared in previous versions of this plan as well as in the RTCA's Core Metrics. As described above, we feel that flying distance is not a good indication of system flexibility in the terminal area.

We have reclassified the fuel usage metrics to the delay/efficiency category from the predictability category, as we feel that this is more appropriate. We also have modified these metrics to focus on fuel usage from the meter fix to the runway threshold rather than simply gate-to-gate fuel burn.

We have changed the previous metric of *variability of acceptance rate less actual arrival rate* to *mean difference of airport acceptance rate less actual arrival rate*. We feel that the mean difference between the actual rate and the AAR is a simpler and better indication of predictability of an airport's throughput than the variability of the difference between the actual rate and the AAR.

We have also added the metric *variability of flight time from the meter fix to the runway threshold*. This is the primary predictability metric for this tool, and its exclusion from previous versions of the Plan was simply an oversight.

We have added the metric *mean actual arrival rate* to the existing metric *mean actual arrival rate for each runway*. We feel that the total arrival rate for the airport is an important metric of access to the NAS.

We have slightly changed the wording of the productivity metric from *distribution of arrivals per runway* to *distribution and throughput of operations per runway/position*. We intend to calculate the distribution of arrivals per runway and the distribution of departures per runway. In addition, we will calculate the average throughput for each position in the terminal airspace.

B.4 CDM Enhanced Ground Delay Program

Seven performance metrics that were previously identified in the FFP1 core list of metrics have been removed as primary metrics. They include

- *Mean taxi-out time*
- *Number of substitutions*
- *Number of diversions*
- *Duration of GDPs*
- *Variation in planned and actual arrival/departure times*
- *Variability of flight distance*
- *Number of airport operations*
- *Number of cancellations*

These metrics are not necessarily flawed but rather, due to problems with measurement and interpretation of results, they have been removed as primary performance metrics. On a limited basis they may still be investigated as possible measures of improved FFP1 performance but will not represented in the principal analysis.

Mean taxi-out time was identified as a measure of delay/efficiency for GDP-E. It has recently been removed as a primary metric because new information is demonstrating that mean taxi-out times are on average *not* impacted by GDP-E. The original theory that some aircraft that are held on the taxi-way during a ground delay program would decrease as a result of GDP-E might have an impact on mean taxi-out times. Several air traffic specialists (including ATCSCC Controllers and ATA) have stated in CDM Working Group meetings that the numbers of aircraft holding on the taxi-ways and the average duration that each aircraft is taxiing is not expected to decrease. Preliminary studies have provided evidence to support this supposition.

The *number of substitutions* was originally chosen as a GDP-E performance metric. Substitution is described as the exchange of arrival slots for certain flights. Simplified substitutions, is described by the CDM Working Group as

> …the need to identify specific pairs of exchanges or substitutions (e.g., flight one is canceled and flight two is substituted into flight one's arrival slot) is eliminated. Users will be allocated a set of arrival slots, and in the initial solution, there will be an initial assignment of flights to slots. If a user cancels or delays a flight that would change the slot assignments, the user simply will report that flight two is now assigned to slot one, flight three is assigned to slot two and so forth. The capability to conduct simplified substitutions is being embedded in the CDM message structure. (Reference 11)

This GDP-E metrics presents the problem of interpretation. It is not known if the change in number of substitutions is good if it increases or decreases. The CDM Working Group is currently investigating the impact of any change in the number of substitutions and will report on the results of the investigation once completed. However, since any potential change in the number of substitutions cannot be read as either positive or negative, this performance metric has been removed from the primary list.

Averted diversions is a major operational impact to airlines and although the *number of diversions* has been included as a GDP-E metric it may be more appropriately used as a Collaborative Routing (CR) metric. Diversions can be obtained using ASQP or ETMS data however the problem of identifying causality will prevent segregating the diversions that would not be effected by GDP-E. As a result, any comprehensive analysis on the number of diversions would require airlines to assist in providing information on causality of diverted flights. Since obtaining the causality information for both the baseline and study periods is questionable, we have removed the number of diversions as a primary metric.

The *duration of GDPs* was once thought to be a promising metric for measuring the impact of GDP-E to ground delay programs. However, due to measurement and normalization problems we have removed it as a primary metric. When examining this metric it would be necessary to normalize for weather and many other contributing factors. We believe that due to the many factors that contribute to the duration of ground delay programs we would end up with small data sets providing little or no confidence in the results. Furthermore, even if various normalization techniques are employed, it is likely that this metric would still be interpreted as being a function of weather.

The *variation in planned and actual arrival/departure times* may change during the use of GDP-E because of the ability of the ATCSCC to better identify airport demand. As before these metrics stem from the belief that GDP-E provides better, more timely information than the procedures that were in use before CDM. Changes in the variation of arrival and departure times indicate the possible contribution that GDP-E has made to NAS users. These predictability measures will likely be estimated as the standard deviation of planned and actual arrival/departure times during GDPs.

Variability of flight distance is an alternative measure to determine significant change between city pairs. It will likely be estimated as the standard deviation of flight distances during ground delay programs. The standard deviation, or the positive square root of the variance, will specifically be used to describe the spread of the distribution. It is possible that given no significant change in the delay/efficiency metric of *mean distance flown* the standard deviation of distance flown may show changes in dispersion. Under this metric, a positive impact to the predictability of the NAS would be seen in a narrower distribution of mean distances flown.

The first metric for discussion is the *number of airport operations.* If a NAS resource (i.e., terminal airspace, enroute airspace) has reached a capacity or throughput constraint during peak traffic levels then improvement in access is generally desirable. As a result an increase in the number of airport operations during peak periods should signify some operational improvement in that resource's capacity. It will be necessary however to determine if GDP-E is the single or partial contributor to the change.

The third access metric is the *number of cancellations*. It is uncertain whether any change in the number of cancellations will take place. Further it is not known if the number of cancellations do change whether an increase or decrease in the number is desired. Some CDM analysts believe that cancellations to GDP airports will fall as a result of GDP-E because of the ability of the airlines to better employ arrival slots. Others believe that the number of cancellations may go up or even stay the same but the characteristics of the cancelled flights will be different. That is, if airlines have more time to determine which flights they want to cancel they may cancel the low revenue flights more often and reallocate the high revenue flight to guarantee an arrival slot and even a reduction in delay. Without the help of the airlines it is uncertain whether the Metrics Team will be able to determine individual characteristics of certain.

B.5 Surface Movement Advisor

We have added the safety metric of *number of operational deviations*, as described previously.

Glossary

AAR	Airport Acceptance Rate
ADOC	Airline Direct Operating Costs
ADR	Airport Departure Rate
AEE	FAA Office of Environment and Energy
ANOVA	Analysis of Variance
AOC	Airline Operations Center
AOZ	FAA Free Flight Phase One Program Office
APO	FAA Office of Aviation Policy and Plans
ARTCC	Air Route Traffic Control Center
ARTS	Automated Radar Terminal System
ASD	FAA Office of System Architecture and Investment Analysis
ASQP	Airline Service Quality Performance
ATC	Air Traffic Control
ATCSCC	Air Traffic Control System Command Center
ATCT	Air Traffic Control Tower
ATL	William B. Hartsfield Atlanta International Airport
ATM	Air Traffic Management
ATS	Air Traffic Services
CAASD	Center for Advanced Aviation System Development
CAEP	Committee on Aviation Environmental Protection
CCLD	Core Capability Limited Deployment
CDM	Collaborative Decision Making
CO	Carbon Monoxide
CO2	Carbon Dioxide
CODAS	Consolidated Operations and Delay Analysis System
CP	Conflict Probe

CR	Collaborative Routing
CTA	Control Time of Arrival
CTAS	Center TRACON Automation System
DART	Data Analysis and Reduction Tool
DFW	Dallas/Fort Worth International Airport
DTW	Detroit Metropolitan Wayne County International Airport
EDCT	Estimated Departure Control Time
ETMS	Enhanced Traffic Management System
EWR	Newark International Airport
FAA	Federal Aviation Administration
FBO	Fixed-Base Operation
FFP1	Free Flight Phase One
FFP1 PO	Free Flight Phase One Program Office
FSM	Flight Schedule Monitor
GA	General Aviation
GDP	Ground Delay Program
GDP-E	Enhanced Ground Delay Program
HC	Hydrocarbons
HCS	Host Computer System
HID	Host Interface Device
ICAO	International Civil Aviation Organization
IDU	Initial Daily Use
IFR	Instrument Flight Rules
IMC	Instrument Meteorological Conditions
IPE	Integrated Predictive Error
LAX	Los Angeles International Airport
LOA	Letter of Agreement

M/A	Monitor Alert
MSP	Minneapolis-St. Paul International Airport
MTR	Monitor, Test, and Recording
N90	New York TRACON
NAS	National Airspace System
NASA	National Aeronautics and Space Administration
NASDAC	National Aviation System Data Analysis Center
NASSI	NAS Status Information
NCDC	National Climatic Data Center
NEXTOR	National Center of Excellence for Aviation Operations Research
NOx	Nitrogen Oxides
NRP	North American Route Program
OAG	Official Airline Guide
ORD	Chicago O'Hare International Airport
PCA	Planned Capability Available
pFAST	Passive Final Approach Spacing Tool
PHL	Philadelphia International Airport
PMP	Program Master Plan
PVT	Passenger Value of Time
RCI	Rate Control Index
RUC	Rapid Update Cycle
RVR	Runway Visual Range
SAR	System Analysis and Recording
SFO	San Francisco International Airport
SMA	Surface Movement Advisor
SOC	System Operations Center
SOP	Standard Operating Procedures

STL	Lambert St. Louis International Airport
SUA	Special Use Airspace
SWAP	Severe Weather Avoidance Program
TFM	Traffic Flow Management
TMA	Traffic Management Advisor
TMA-SC	Traffic Management Advisor – Single Center
TMC	Traffic Management Coordinator
TRACON	Terminal Radar Approach Control
URET	User Request Evaluation Tool
VMC	Visual Meteorological Conditions
ZAU	Chicago ARTCC
ZBW	Boston ARTCC
ZDC	Washington DC ARTCC
ZDV	Denver ARTCC
ZFW	Fort Worth ARTCC
ZID	Indianapolis ARTCC
ZKC	Kansas City ARTCC
ZLA	Los Angeles ARTCC
ZMA	Miami ARTCC
ZME	Memphis ARTCC
ZMP	Minneapolis ARTCC
ZNY	New York ARTCC
ZOA	Oakland ARTCC
ZOB	Cleveland ARTCC
ZTL	Atlanta ARTC